

INJURY SEVERITY OF MULTI-VEHICLE CRASH IN RAINY WEATHER

By

Soyoung Jung

(Corresponding Author)*

Graduate Researcher

University of Wisconsin-Madison

Department of Civil and Environmental Engineering

1241 Engineering Hall

1415 Engineering Drive

Madison, WI 53706

Phone: (608) 262-2524

Email: jung2@wisc.edu

Xiao Qin, Ph.D., P.E.

Assistant Professor

South Dakota State University

Department of Civil and Environmental Engineering

148 Crothers Engineering Hall

Brookings, SD 57007

Phone: (605) 688-6355; Fax: (605) 655-6474

Email: Xiao.Qin@sdstate.edu

David A. Noyce, Ph.D., P.E.

Associate Professor

University of Wisconsin-Madison

Department of Civil and Environmental Engineering

1415 Engineering Drive

Madison, WI 53706

Phone: (608) 265-1882; Fax: (608) 262-5199

Email: noyce@engr.wisc.edu

Submission Date:

August 1st, 2009

Length of Paper:

5419 words, 5 tables and 1 figures @ 250 words each

6919 equivalent words

ABSTRACT

As part of the Wisconsin road weather safety initiative, the objective of this study is to microscopically assess rainy weather effect on the severities of multi-vehicle involved crashes on Wisconsin interstate highways utilizing a polychotomous response model, sequential logistic regression.

Weather related factors considered in this study included estimated rainfall intensity for 15 minutes prior to a crash occurrence, water film depth, temperature, wind speed/direction, stopping sight distance and car-following distance at the crash moment. For each crash observation, weather station data around the crash location were interpolated using the inverse squared distance method. Non-weather factors such as road geometries, traffic conditions, collision manners, vehicle types, and driver and temporal attributes were also considered. The sequential logistic regression was tested with forward and backward formats for the polychotomous outcomes of multi-vehicle crash severity. The best format to predict the multi-vehicle crash severities in rainy weather was selected by combining measures of model performance for goodness of fit, parameter significance, and prediction accuracies.

In conclusion, the backward sequential logistic regression model produced the best results for predicting crash severities in rainy weather where water film depth, the number of traffic lanes, tangent roadway section, peak traffic hours, at-fault driver's action at the crash moment, standard deviation of 5-minute traffic volume and safety belt usage were found to be statistically significant. These findings can be used to determine the probabilities of multi-vehicle crash severity in rainy weather and provide quantitative support on improving road weather safety via weather warning systems, highway facility improvements, and traffic management.

Keywords: Road weather safety, rainy weather effect, multi-vehicle crash severities, sequential logistic regression

INTRODUCTION

According to Wisconsin Traffic Crash Facts (1) from the Wisconsin Department of Transportation (WisDOT), there are 3,296 injury and fatal crashes in rainy weather, the greatest number of all kinds of inclement weather conditions from 1999 to 2006. Specifically on Wisconsin interstate highways in rainy weather, multi-vehicle involved crashes occurred more frequently than single-vehicle crashes. From 1999 to 2006, 899 multi-vehicle crashes occurred on Wisconsin interstate highways in rainy weather and the number of multi-vehicle crashes is approximately 1.5 times more than the number of single-vehicle crashes (2). Even though multi-vehicle crashes occurred more frequently on Wisconsin interstate highways in rainfall, reduction of severe multi-vehicle crashes is a way to improve road weather safety because the severe crashes involving injuries or fatalities cause much more economic loss than property damage only crashes.

Weather has been frequently cited and found as one of the factors contributing to either a more or less severe crash. The approaches for injury severity prediction with the weather factors vary from one to another, depending on the purpose of study and data availability. Based on the purpose of this study, the focus of literature review will be specifically on the factors caused by rain at the time of multi-vehicle involved crashes.

First, rain-derived factors can affect multi-vehicle crash severities differently by collision types. Shankar et al. estimated a nested logit model of accident severity that occurred on Washington rural interstate (13). In their study, wet-pavement rear-end accident indicator was found to increase the likelihood of possible injuries, capturing the effect of rear-end accidents occurring in rainy weather. They explain the reason that rainy weather conditions make vehicles in front more difficult to see and increase the distance required to stop, resulting in the injury crashes. Duncan et al. used ordered probit model to identify specific variables significantly influencing levels of injury in truck-passenger car rear-end involvements on divided roadways (14). Interaction wet and grade was found to significantly increase all injury propensities in their study. Similarly, Yan et al.'s study showed that the wet and slippery road surface greatly contributed to rear-end accidents at signalized intersections compared to a dry road surface (15).

Special attention has been given to vehicle types for multi-vehicle crash severity prediction. In the study by Duncan et al., wet grades was found to increase particularly severe injuries to passenger car occupants in truck-passenger car rear-end crashes (14). Haquea et al. utilized binary logit model to differentiate between at-fault and not-at-fault cases to identify the factors that contribute to the fault of motorcyclists involved in crashes (16). In their study, motorcyclists were more likely to be victims than at-fault in multi-vehicle crashes and wet road surface was found to increase the likelihood of at-fault crashes at non-intersections.

Rain-related effect on multi-vehicle crash severities has been identified along with roadway characteristics as well. Khorashadi et al. explored the differences between urban and rural driver injuries in large truck involved accidents using multinomial logit analysis (17). In their study, raining was found to increase the likelihood of complaint of pain accidents only in urban area. Deng et al. predicted the severities of head-on crashes that occurred on Connecticut two-lane roads utilizing ordered probit model (18). 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 the likelihood of injury severity associated with weather conditions. Hill and Boyle investigated fatality and incapacitating injuries to occupants of passenger vehicles by a logistic regression model (19). The 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.

Though numerous studies have been conducted in hopes of identifying the contributing factors to crash severities (3, 4, 5, 6, 7, 8, 9, 10, 11, 12), in these crash severity-based studies, weather is just one of the contributing factors, not necessarily the focus. This study will identify a variety of significant predictors that contribute more serious multi-vehicle crash consequences particularly using only rainy weather related crashes. Research findings from this study will provide guidance on countermeasures to prevent severe crashes related to rainy weather and improve overall safety.

SEQUENTIAL LOGISTIC REGRESSION MODEL

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 dataset varies from one to another due to their limitations. Crash severities are not only multiple discrete outcomes but also inherently ordered. However, the multinomial and nested logit models do not account for the ordering of crash severities (20, 21, 22). 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. In reality, it is too arbitrary to assume that all coefficients across the response outcomes are the same (21, 22).

Alternatively, a generalized version of the ordered logit model was used to relax the proportional odds assumption (23). However, the generalized ordered response model with separate parameter coefficients across the ordered response levels is recommended only to conclude that the proportional odds assumption is valid because the model is very anti-conservative (24). Based on the purpose of this study, a variety in sets of predictors across various severity levels is one of the most important issues. 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 this study, consequently, sequential logistic regression approach was selected to predict rainy weather crash severities because this method not only accounts for the inherent ordering of crash severities but also allows different sets of regression parameters across the severity levels.

The sequential logistic regression is composed of a series of standard logistic regression. Based on 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 (25):

$$\frac{P(X)}{1 - P(X)} = \text{EXP}(\alpha + \beta X) \quad (1)$$

Where,

$P(X) = P(Y = y | X)$ = Probability of response outcome,

Y = Response variable,

$y = 0$ or 1 ,

α = Intercept parameter,

β = Vector of parameter estimate,

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 one-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 sub-sample is used after removing observations of a certain crash severity used in the previous stages (26). In order 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 1: Property damage only (PDO) vs. Others

Stage 2: Non-incapacitating/possible injuries vs. Fatal/incapacitating injuries

Backward format:

Stage 1: Fatal/incapacitating injuries vs. Others

Stage 2: Non-incapacitating/possible injuries vs. PDO

Using the standard logistic regression concept at each stage in two formats, the probabilities of crash severity levels can be written as follows:

Forward format:

$$\text{Stage 1: } \frac{1-P_1}{P_1} = \text{EXP}(\alpha_1 + \beta X_1) = h_1 \quad (2)$$

$$\text{Stage 2: } \frac{P_3}{P_2} = \text{EXP}(\alpha_2 + \beta X_2) = h_2 \quad (3)$$

$$P_1 = \frac{1}{1+h_1} \quad (4)$$

$$P_2 = \frac{h_1}{(1+h_1)(1+h_2)} \quad (5)$$

$$P_3 = \frac{h_1 h_2}{(1+h_1)(1+h_2)} \quad (6)$$

Backward format:

$$\text{Stage 1: } \frac{P_3}{1-P_3} = \text{EXP}(\alpha_1 + \beta X_1) = I_1 \quad (7)$$

$$\text{Stage 2: } \frac{P_2}{P_1} = \text{EXP}(\alpha_2 + \beta X_2) = I_2 \quad (8)$$

$$P_1 = \frac{1}{(1+I_1)(1+I_2)} \quad (9)$$

$$P_2 = \frac{I_2}{(1+I_1)(1+I_2)} \quad (10)$$

$$P3 = \frac{I1}{1 + I1} \quad (11)$$

Where,

P1 = Probability of PDO,

P2 = Probability of non-incapacitating/possible injury,

P3 = Probability of fatal/incapacitating injury.

MODEL PERFORMANCE MEASURES

Typical measures of model performance for goodness of fit and prediction accuracy are likelihood ratio test and prediction accuracy classification, respectively. These measures are synthetically considered to assess crash severity prediction models in this study.

The likelihood ratio (LR) test reveals whether or not global null hypothesis for a specific model is rejected. In other words, an estimated model containing at least one non-zero parameter coefficient is better fit than constant only model when p-value of LR test is less than a conventional criterion.

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 cut-point. Otherwise, it is classified as a non-event. In the statistical term, the rate of actual events that are also predicted to be events is called sensitivity. Similarly, the rate of actual non-events that are also predicted to be non-events is called specificity. The overall predictive power of a model depends on the proportion 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 rate. The false positive rate is the ratio of the number of non-events incorrectly classified as events to the total classified events while the false negative rate is a ratio of the number of events incorrectly classified as non-events to the sum of total classified non-events.

Even though the predictive power of a model can be measured for all severity levels, sensitivity and false negative rate are emphasized for the highest crash severities (fatal and incapacitating injuries) because of their enormous economic loss. Hence, in this study, a model that produces high sensitivity and low false negative rate at the classification stage for fatal and incapacitating injuries is considered as a good one.

Note that the prediction accuracy in a model are variant by a probability cut-point since the model classifies an observation based on the given probability cut-point. More severe crashes (event) do not frequently occur compared with less severe crashes (non-event). From this perspective, the probability cut-point may be determined by practical consideration. Since desirable prediction models should fit the field data well, probability cut-points used in this study are determined by overall trends of actual event proportions in the field data.

DATA COLLECTION AND PROCESSING

The study area consisted of approximately 75 miles of Southeastern Wisconsin highway segments including I-43, I-94, I-43/94 and I-43/894, where rainy weather crash frequency, average annual daily traffic and vehicle miles traveled was higher than any other interstate highway segments between 2004 and 2008. The study area is shown in Figure 1.



FIGURE 1 Study area (27).

Data Source

Crash dataset for multi-vehicle crashes occurring in rainy weather were obtained from the 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 form data homogeneity: wet pavement, multi-vehicle included in a crash, interstate highway divided by barrier, no construction zone, no hit and run and no pedestrian involved in a crash. Consequently, 536 crashes were produced in the study area from 2004 to 2008. Crash dataset included variables indicating severity, roadway geometries, driver demographics, collision types, vehicle types, pavement conditions, and temporal and weather information.

Incapacitating injury (type A) and fatal (type K) crashes were combined as the highest level of crash severity to obtain a meaningful sample size (28). Possible injury (type C) and non-incapacitating injury (type B) crashes were combined as the second highest level of crash severity because they were not distinguishable. Property Damage Only crashes made up the lowest level of crash severity. Crash frequency and the SAS program coding by the severities are provided in Table 1.

TABLE 1 Multi-Vehicle Crash Frequency in Rainy Weather

Injury Severity	Sequential Logistic Regression			
	Forward Format		Backward Format	
	Stage 1	Stage 2	Stage 1	Stage 2
Fatal and Incapacitating Injury	13 (1) ¹	13 (1)	13 (1)	-
Non-incapacitating and Possible Injury	187 (1)	187 (0)	187 (0)	187 (1)
PDO	336 (0)	-	336 (0)	336 (0)
Total	536	200	536	523

¹ SAS coding of crash severity level

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

V-SPOC traffic detectors collect and archive traffic data in Southeast Wisconsin every 30 seconds. Average vehicle volume, speed and occupancy data for five-minute intervals were obtained for one hour prior to each crash. The associated standard deviations providing one hour temporal buffers prior to the crash were also collected due to the difference in density between crash and detector locations in study area.

Based on the objective of this study, one of the most important tasks was to collect microscopic weather data at time of crash. However, there were few weather data sources to provide minute base measurements. A website run by Weather Underground Incorporation, combined with local weather data, delivered the most reliable and real-time weather data for Wisconsin (29). Based on these data, 6 airport weather stations and 10 private weather stations were considered to obtain microscopic weather data in this study.

Weather Parameters

Weather data directly collected from a weather station were temperature, wind speed/direction, rainfall precipitation and rainfall duration. To reflect real-time weather conditions at the crash moment, weather data was estimated by interpolating between nearest three weather stations because weather data such as rainfall intensity or wind speed show geometrical and temporal variety. Water film depth, stopping sight distance (SSD), and deficiency of car-following distance (DCD) were estimated by hourly rainfall precipitation, traffic, and road geometry data.

Rainfall Intensity

Rainfall intensity is defined as the rainfall precipitation divided by measurement interval. The rainfall intensity reflects visibility on highway in rainy weather conditions. Using three weather station data for each crash location, the average measurement interval of rainfall precipitation was 15 minutes. Therefore, rainfall precipitation for 15 minutes prior to a crash was adopted as the real-time rainfall intensity at the crash moment. Compared to the weather data measurement intervals mentioned in the previous studies, 15-minute measurement interval used in this study was a more microscopic reflection of the real-time rainfall intensity at a crash moment.

Water Film Depth

The water film created by rainfall exists between the tire and the pavement surface, leading to a decrease in skid resistance. Russam and Ross gave the following empirical method to estimate water film depth (30):

$$D = 0.046 \frac{(WS'I)^{1/2}}{S^{1/5}} \quad (12)$$

$$S = (S_l^2 + S_c^2)^{1/2} \quad (13)$$

Where,

D = Water film depth (mm/hr)

I = Rainfall intensity (mm/hr)

S' = S/S_c

S_l = Longitudinal slope (%)

S_c = Slope of pavement cross section (%)

W = Width of pavement (m)

Stopping Sight Distance and Deficiency of Car-following Distance

In this study, there were no direct visibility data for highways. Therefore, stopping sight distance (SSD) and deficiency of car-following distance (DCD) were considered as the surrogate measures for highway visibility at the time of the crash. First of all, SSD formula is as follows (31).

$$SSD = 1.47Vt + 1.075 \frac{V^2}{a} \quad (14)$$

Where,

V = vehicle speed (mi/hr)

t = brake reaction time (sec)

a = deceleration rate (ft/s²)

According to a detailed study about pavement conditions, the coefficient of wet pavement friction is associated with water film depth and vehicle speed (32). The study shows the relations among friction force, vehicle speed and water film depth.

Combining the relation in the study and using pavement surface material information from Wisconsin STN highway log, a deceleration rate value to apply to the SSD equation was obtained by correlating to the pavement friction coefficient (33). In addition, 2.5 seconds exceeding the 90th percentile of reaction time for all drivers were used for brake reaction time to encompass the capabilities of most drivers (31). Consequently, SSD was calculated by the maximum wet pavement deceleration rate, vehicle speed from traffic detector data, and brake reaction time.

Strictly speaking, vehicle speed in SSD formula should be individual vehicular speeds, so is the gap between every pair of cars. In this study, the average of five-minute traffic detector data containing the crash occurrence time was used to surrogate the real-time prevailing traffic conditions at the crash moment.

DCD represents the risk of losing control caused by driver overcorrection for avoiding any potential conflict. DCD is calculated by the following formula.

$$DCD = SSD - AVG \quad (15)$$

Where,

SSD = stopping sight distance

AVG = average vehicle gap

In DCD formula, AVG is obtained by subtracting average vehicle length from inverse of vehicle density calculated by traffic detector data (34).

As a result of data collection from several data sources, explanatory variables and the associated category coding used in this study are shown in Table 2.

TABLE 2 Explanatory Variables Used in Crash Severity Prediction Model

Variable	Min.	Max.	Mean	Category Coding
At-fault driver's sex	-	-	-	Female=1, Male=2
Alcohol or drug	-	-	-	Sobriety=1, Under alcohol/drug effect=2
Safety belt	-	-	-	Use of safety belt=1, Non-used =2
At-fault driver's action	-	-	-	Going straight=1, Lane change/merging/overtaking=2, Negotiating curve=3, Slowing or stopped=4
Curve direction	-	-	-	Curve to the right=1, Curve to the left=2
Injury transport	-	-	-	Injured people transported to hospital=1, Others=2
Terrain	-	-	-	Horizontal curve=1, Vertical curve=2, Horizontal/vertical curve=3, Tangent/flat=4
First harmful spot	-	-	-	Ramp/gore=1, Shoulder/outside shoulder=2, Median=3, On roadway=4
Pavement surface	-	-	-	Asphaltic cement plant mix/rigid base=1, Others=2
Lighting condition	-	-	-	Daylight=1, Dusk/Dawn/Dark=2, Night but street light=3
Crash type	-	-	-	Median related=1, Non-collision=2, Fixed object=3
First harmful collision	-	-	-	Sidewipe=1, Rear-end=2, Others=3
At-fault driver's vehicle	-	-	-	Car=1, Truck(straight)/truck-tractor=2, Motor cycle=3
Time of day	-	-	-	Peak-hour (6-8 a.m. & 3-5 p.m.)=1, Off-peak=2
Day of week	-	-	-	Tuesday to Thursday=1, Monday/Friday=2, Saturday/Sunday= 3
Quarter of year	-	-	-	December to February=1, March to May=2, June to August=3, September. to November=4
Wind direction	-	-	-	No wind=1, North=2, East=3, South=4, West=5
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 ¹	13/16	7/11	-
Speed limit (mi/h)	35	65	55	-
Avg. 5-min V ²	5	172	94	-
Avg. 5-min SPD ³	1	91	48	-
Avg. 5-min O ⁴ (%)	0.45	49.11	13.00	-
S.D ⁵ . of V	0.94	69.46	9.78	-

S.D. of SPD	0.32	63.29	5.88	-
S.D. of O	0.10	18.23	3.02	-
SSD (ft)	5	4509	384	<238=1, [238, 494]=2, >494=3
DCD (ft)	0	3825	153	<55 =1, [55, 225]=2, 225> =3
Wind speed (km/h)	0.0	43.9	9.0	<2.6=1, [2.6, 13.5]=2, >13.5=3
Temperature (°C)	0.1	29.4	11.0	<5.0 =1, [5.0, 17.0]=2, 17.0>=3
Water film (mm/h)	0.00	0.45	0.06	<0.02=1, [0.02, 0.10]=2, >0.1=3
RI ⁶ (mm/15min)	0.00	8.96	0.26	<0.05 =1, [0.05, 0.3]=2, 0.3>=3

¹ Left shoulder width/Right shoulder width

² Volume

³ Vehicle speed

⁴ Occupancy

⁵ Standard deviation

⁶ Rainfall intensity

Weather Data Interpolation

To estimate unknown weather data, a study regarding the comparison of interpolation methods concluded that the inverse squared distance method is stable and appropriate for the localized field with short spatial correlation length scale and large variability (35). The minimum number of weather stations to apply the inverse squared distance interpolation was three (36). Therefore, the inverse squared distance interpolation was utilized to estimate localized weather data at the crash moment.

Due to the geographical and temporal variety of data, rainfall intensity during 15 minutes prior to a crash, water film depth and wind speed were interpolated between three weather stations nearest to the crash spot by the inverse squared distance method. However, temperature data from one weather station nearest to each crash was used due to its proximity.

ANALYSIS AND DISCUSSION

In this study, PROC LOGISTIC statement in SAS 9.1 was used to estimate sequential logistic regression models on the basis of rain-related multi-vehicle crashes with a significance level of 0.10 for retaining predictors in the models. The modeling process was as follows.

First of all, bivariate logistic regression of each explanatory variable was performed to choose an individual predictor which correlated to crash severity. Since weather effect on crash severity is the primary interest of this study, the continuous weather data were specifically transformed to a categorical variable by quantile if any continuous weather variable was not selected by the bivariate logistic regression.

Next, correlation between predictors selected by the bivariate logistic regression was identified by Pearson's correlation coefficient or likelihood ratio chi-squared test in order not to omit significant predictors in multiple logistic regression models. After the correlation test, several combinations with the maximum number of uncorrelated predictors were constructed for the next step.

Then, stepwise variable selection was conducted to select a multiple logistic regression model for each combination of the uncorrelated predictors. Comparing the multiple logistic regressions by goodness of fit, parameter significance and prediction accuracy, the best multiple logistic regression was chosen for each format of sequential logistic regression. Using on the same

measures of model performance, one of the two formats was finally selected to predict multi-vehicle crash severities in rainy weather. Based on proportions of actual more severe crashes (event) at each stage of each format, the range for event probability cut-points is provided in the following Table.

TABLE 3 Event Proportion in Study Area

Year	Forward Format		Backward Format	
	Stage 1	Stage 2	Stage 1	Stage 2
2004	0.34	0.05	0.02	0.33
2005	0.37	0.13	0.05	0.34
2006	0.33	0.07	0.02	0.31
2007	0.32	0.07	0.02	0.31
2008	0.37	0.08	0.03	0.35
Minimum	0.32	0.05	0.02	0.31
Maximum	0.37	0.13	0.05	0.35
Average	0.35	0.08	0.03	0.33

Forward Format

In Table 4, the global null hypothesis was rejected at both stages, indicating the estimated models were better to predict crash severities than constant only model and all of the parameter estimates for explanatory variables were significant. At stage one, changing lanes or merging into traffic (driver action 2) and slowing/stopping (driver action 4) by at-fault driver decreased the likelihood of injury crashes in the rainfall while negotiating curve (driver action 3) by at-fault driver increased the likelihood. The effects of at-fault driver's actions at the crash moment imply that driver's caution in the rainfall works comparatively less on curves. Interestingly, large standard deviation of 5-minute traffic volume (SDV) was found to decrease the likelihood of injury crashes slightly, which may not be consistent to general expectation. The reason may also be that a large variety of traffic volume in rainy weather leads to more cautious driving.

At the second stage, vertical curves (terrain 2) and horizontal/vertical curves (terrain 3) were much stronger than other factors to increase the likelihood of severe crashes with fatal and incapacitating injuries in the rainfall while more travel lanes decreased the likelihood of the severe crashes. It was found that deficiency of car-following distance (DCD) and peak traffic hour were likely to decrease the most severe crashes implying rainfall effects for DCD and peak hours at the second stage are similar to the effect for SDV at the first stage.

Within the range of probability cut-points shown in classification table, overall prediction accuracy and sensitivity were reasonable at both stages. Especially, overall prediction accuracies at stage two was much higher than those at stage one. However, sensitivities at the second stage showed a variety within the range of probability cut-points, indicating the power of forward format is not stable to predict higher severities.

TABLE 4 Forward Format of Multiple Sequential Logistic Regression

Stage 1			
LR test	ChiSq	D.F.	Pr > ChiSq
	24.0870	4	<.0001

	Parameter	Estimate	S.E.	Odds Ratio	Pr > ChiSq	
Analysis of maximum likelihood estimates	Intercept 1	0.0483	0.0592	-	0.1986	
	Driver action 2	-0.8546	0.3026	0.425	0.0047	
	Driver action 3	0.7685	0.4381	2.156	0.0794	
	Driver action 4	-0.5278	0.2058	0.590	0.0103	
	SDV	-0.0345	0.0158	0.966	0.0292	
Classification table	P _{cut-off}	Overall	Sensitivity	Specificity	FP	FN
	0.32	49 %	72 %	45 %	60 %	32 %
	0.33	52 %	70 %	42 %	58 %	30 %
	0.34	56 %	67 %	50 %	56 %	28 %
	0.35	58 %	63 %	55 %	55 %	29 %
	0.36	59 %	62 %	58 %	53 %	28 %
	0.37	60 %	60 %	59 %	53 %	28 %
Stage 2						
LR test	ChiSq	D.F.	Pr > ChiSq			
	30.6945	5	<.0001			
Analysis of maximum likelihood estimates	Parameter	Estimate	S.E.	Odds Ratio	Pr > ChiSq	
	Intercept	3.3406	1.9701	-	0.0900	
	Terrain 2	3.2543	0.8879	25.902	0.0002	
	Terrain 3	2.8603	1.3177	17.466	0.0300	
	Peak hour	-1.8248	0.8666	0.161	0.0352	
	DCD	-0.0093	0.0038	0.991	0.0130	
Lane number	-1.8159	0.6939	0.163	0.0089		
Classification table	P _{cut-off}	Overall	Sensitivity	Specificity	FP	FN
	0.05	72 %	77 %	72 %	84 %	2 %
	0.07	76 %	70 %	77 %	83 %	3 %
	0.09	80 %	54 %	81 %	83 %	4 %
	0.11	80 %	46 %	83 %	86 %	5 %
	0.13	89 %	31 %	93 %	77 %	5 %

Backward Format

According to Table 5, small P-values in LR test at both stages reveals that the selected models with significant explanatory variables is better fit than the global null models.

TABLE 5 Backward Format of Multiple Sequential Logistic Regression Model

Stage 1			
LR test	ChiSq	D.F.	Pr > ChiSq
	29.7022	4	<.0001

	Parameter	Estimate	S.E.	Odds Ratio	Pr > ChiSq	
Analysis of maximum likelihood estimates	Intercept	2.7670	1.7571	-	0.1153	
	Terrain 4	-2.1498	0.6196	0.117	0.0005	
	Peak hour	-1.7829	0.7921	0.168	0.0244	
	Water film depth	-17.729	8.5817	0.001	0.0388	
	Lane number	-1.2871	0.5660	0.276	0.0230	
Classification table	P_{cut-off}	Overall	Sensitivity	Specificity	FP	FN
	0.02	74 %	69 %	74 %	93 %	1 %
	0.03	80 %	69 %	80 %	92 %	1 %
	0.04	90 %	69 %	90 %	85 %	1 %
	0.05	91 %	69 %	91 %	83 %	1 %
Stage 2						
LR test	ChiSq	D.F.	Pr > ChiSq			
	31.2532	5	<.0001			
	Parameter	Estimate	S.E.	Odds Ratio	Pr > ChiSq	
Analysis of maximum likelihood estimates	Intercept	1.5969	0.7149	-	0.0255	
	Driver action 2	-1.0987	0.3364	0.333	0.0011	
	Driver action 3	0.8176	0.4443	2.265	0.0657	
	Driver action 4	-0.4482	0.2092	0.639	0.0321	
	Safety belt	-1.6850	0.7059	0.185	0.0170	
	SDV	-0.0315	0.0160	0.969	0.0484	
Classification table	P_{cut-off}	Overall	Sensitivity	Specificity	FP	FN
	0.31	48 %	74 %	34 %	62 %	29 %
	0.32	52 %	71 %	42 %	60 %	27 %
	0.33	56 %	68 %	49 %	58 %	26 %
	0.34	58 %	64 %	55 %	56 %	26 %
	0.35	60 %	62 %	58 %	55 %	26 %

At stage one, the odds ratio for water film depth was much less than 1. That is, fatal and incapacitating injury crashes decreased as water film depth increased. This result implies that drivers tend to recognize the risk of rain covered pavements and adjust driving behavior accordingly. Straight roadway section (terrain 4) and more travel lanes decreased the likelihood of the most severe crashes in rainy weather conditions. The odds ratio for peak hour was also less than 1, which is the same result as the second stage of forward format.

At stage two, wearing a safety belt was found to decrease non-incapacitating and possible injury crashes. Except for the safety belt, selected predictors and their effects on crash severities at the second stage were equivalent to those at the first stage of forward format.

Based on classification tables, overall prediction accuracies and sensitivities within the range of probability cut-points were reasonable at both stages. In particular, overall prediction accuracies at the first stage of backward format were higher than those at the second stage of forward format. Moreover, consistently high sensitivities and low false negative rates (FN) at the second stage imply that backward format is the most desirable to predict the most severe crashes.

Comparing multiple regression results based on goodness of fit, parameter significance and prediction accuracies, the backward sequential logistic regression model outperforms forward sequential logistic regression model in predicting multi-vehicle crash severity levels in rainy weather. Specifically, the backward sequential logistic regression model was found to be effective to predict the highest level of crash severity. In addition, weather-related factor such as water film depth was explicitly and significantly identified in the backward format.

Therefore, the backward format of sequential logistic regression model was selected as the best final model for predicting multi-vehicle crash severity that occurred on high-speed highways in rainy weather.

CONCLUSIONS AND FUTURE RESEARCH

In previous studies, the rainy weather-related factors lacked the accuracy and sophistication to reflect real-time pavement surface conditions and visibility during rainfall. 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 more microscopic weather data, this study assessed rainfall effects on the severities of multi-vehicle 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-minute rainfall intensity, water film depth, stopping sight distance, and deficiency of car-following distance that have not been frequently considered in the previous studies. In addition, estimated or measured weather factors were interpolated between three weather stations by inverse squared interpolation method for each crash location.

In this study, sequential logistic regression models were applied to predict polychotomous response outcomes such as crash severities because the sequential logistic regression is more flexible to reflect variant predictor effects on the response categories. The sequential logistic regression models were further divided into the forward format from the lowest injury severity to the highest one and the backward format reversing the sequence.

As a result, the backward format of sequential logistic regression model outperformed the forward format in predicting crash severity levels, especially fatal and incapacitating injuries, with the higher prediction accuracy and it significantly identified the effect of water film depth derived by rain precipitation. In the backward sequential logistic regression model, following variables were significantly identified: water film depth, the number of travel lanes, tangent roadway section, peak hour, changing lanes/merging /overtaking, slowing/stopping, standard deviation of 5-minute traffic volume and safety belt usage. Note that water film depth, lane change/merging/overtaking prior to the crash time, variant traffic volume and peak hour decreased the likelihood of more severe crashes, which implies that rainfall may affect defensive driving more than expected.

Thus, the backward sequential logistic regression model is considered to be the most appropriate for determining the probability of multi-vehicle crash severity in rainy weather. The resultant findings in this study can be used to provide quantitative support on improving road weather safety via weather warning systems, highway facility improvements, and traffic management.

REFERENCES

1. *Wisconsin Traffic Crash Facts*. Wisconsin Department of Transportation, 1999-2006.
2. *MV4000 Crash Database Query Tools*. WisTransPortal Project Homepage. <http://transportal.cee.wisc.edu/> Accessed Jan. 2009.
3. Abdel-Aty, M., and Pemmanaboina, R. Calibrating a Real-Time Traffic Crash Prediction Model Using Archived Weather and ITS Traffic Data, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, 2006, pp. 167-174.
4. Dissanayake, S., and Lu, J. Analysis of Severity of Young Driver Crashes: Sequential Binary Logistic Regression Modeling, In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, TRB, National Research Council, Washington, D.C., 2002, pp. 108-114.
5. Donnell, E., and Mason, J. Jr. Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1897, TRB, National Research Council, Washington, D.C., 2004, pp. 55-63.
6. Golob, T., and Recker, W. Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions, *Journal of Transportation Engineering*, Vol. 129, 2003, pp. 342-353.
7. Khan, G., Qin, X., Noyce, D. Spatial Analysis of Weather Crash Patterns, *Journal of Transportation Engineering*, Vol. 134, 2008, pp. 191-202.
8. Qin, X., Noyce, D., Lee, C. Snowstorm Event-Based Crash Analysis, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1948, TRB, National Research Council, Washington, D.C., 2006, pp. 135-141.
9. Savolainen, P., and Tarko, A. Safety Impacts at Intersections on Curved Segments, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1908, TRB, National Research Council, Washington, D.C., 2005, pp. 130-140.
10. Shankar, V., Fred Mannering, F., and Barfield, W. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies, *Accident Analysis and Prevention*, Vol. 27, 1995, pp. 371-389.
11. Yau, K.. Risk Factors Affecting the Severity of Single Vehicle Traffic Accidents in Hong Kong, *Accident Analysis and Prevention*, Vol. 36, 2004, pp. 333-340.
12. Yau, K., Lo, P., and Fung, S.. Multiple-Vehicle Traffic Accidents in Hong Kong, *Accident Analysis and Prevention*, Vol. 38, 2006, pp. 1157-1161.
13. Shankar, V., Mannering, F., and Barfield, W. Statistical Analysis of Accident Severity on Rural Freeways, *Accident Analysis and Prevention*, Vol. 28, 1996, pp. 391-401.
14. Duncan, C., Khattak, A., and Council, F. Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1635, TRB, National Research Council, Washington, D.C., 1998, pp. 63-71.
15. Yan, X., Radwan, E., and Abdel-Aty, M. Characteristics of Rear-End Accidents at Signalized Intersections Using Multiple Logistic Regression Model, *Accident Analysis and Prevention*, Vol. 37, 2005, pp. 983-995.
16. Haquea, M., Chinb, H., and Huang, H. Modeling Fault Among Motorcyclists Involved in Crashes, *Accident Analysis and Prevention*, Vol. 41, 2009, pp. 327-335.
17. Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F. Differences in Rural and Urban Driver-Injury Severities in Accidents Involving Large-Trucks: An Exploratory Analysis, *Accident Analysis and Prevention*, Vol. 37, 200, pp. 910-921.

18. Deng, Z., Ivan, J., and Garder, P. Analysis of Factors Affecting the Severity of Head-On Crashes Two-Lane Rural Highways in Connecticut, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1953, TRB, National Research Council, Washington, D.C., 2006, pp. 137-146.
19. Hill, J., and Boyle, L. Assessing the Relative Risk of Severe Injury in Automotive Crashes for Older Female Occupants, *Accident Analysis and Prevention*, Vol. 38, 2006, pp. 148-154.
20. Abdel-Aty, M., Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models, *Journal of Safety Research*, Vol. 34, 2003, pp. 597-603.
21. Milton, J., Shankar, V., Mannering, F. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis, *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 260-266.
22. Wang, X, and Abdel-Aty, M. Analysis of Left-turn Crash Injury Severity by Conflicting Pattern Using Partial Proportional Odds Models, *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1674-1682.
23. Eluru, N., Bhat, C., and Hensher, D. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes, *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1033-1054.
24. Peterson, B., Harrell, F.E. Jr. Partial Proportional Odds Model for Ordinal Response Variables, *Applied Statistics*, Vol. 39, 1990, pp. 205-217.
25. Kleinbaum, D., and Klein, M. *Logistic Regression: A Self-Learning Text*, Springer, New York, 2002.
26. Maddala, G. *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York, 1983.
27. *Maps*, Google. <http://maps.google.com/maps?hl=en&tab=wl>. Accessed Jul. 2009.
28. Agresti, A. *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc, New York, 1996.
29. *Wisconsin Weather Station Data*, Weather Underground Incorporation Homepage. <http://www.wunderground.com/US/WI/>. Accessed Jul. 2009.
30. Russam, K. and Ross, N. *The Depth of Rain Water on Road Surfaces*, Ministry of Transport Report No. LR 236, Road Research Laboratory, 1968.
31. American Association of State Highway and Transportation Officials (AASHTO). *A policy on Geometric Design of Highways and Streets*. AASHTO, 2004.
32. Kokkalis, A., and Panagouli, O. Factual Evaluation of Pavement Skid Resistance Variation: Surface Wetting, *Chaos, Chaos, Solitons and Fractals*, Vol. 9, 1998, pp. 1875-1890.
33. *Reference Tables-Coefficient of Friction*, Engineer's Handbook Homepage. <http://www.engineershandbook.com/Tables/frictioncoefficients.htm>. Accessed Jul. 2009.
34. Roess, R., Prassas, E., and Mcshane, W. *Traffic Engineering*, Pearson Education, Inc., New Jersey, 2004.
35. Patrick, N., and Stephenson, D. Spatial Variation of Rainfall Intensities for Short Duration Storms, *Hydrological Sciences*, Vol. 35, 1990, pp. 667-680.
36. Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, 2007.

