Identification of Factors Leading to High Severity of Crashes in Rural Areas Using Ordered Probit Modeling

by Sunanda Dissanayake and Indike Ratnayake

This study made an effort to identify critical factors contributing to increased crash severities on rural highways. Crash data from the Kansas Accident Reporting System (KARS) database was analyzed and crash severity was modeled using ordered choice models. Many driver-related factors, such as alcohol involvement, lack of seat belt usage, excessive speed, and driver ejections because of the crash contribute to the increased severity of crashes in rural areas. Also, severities of single-vehicle crashes are higher than two-vehicle and animal-vehicle crashes. Factors related to roadway geometry such as sharp curves and steep grades are also found to contribute to the increased crash severity in rural areas.

INTRODUCTION

A total of 42,815 people died because of highway crashes in the United States in 2002 (NHTSA 2002). About 60% of those fatalities occurred on rural highways which account for more than 75% of the total highway mileage in the United States. However, total vehicle miles traveled on rural highways accounted only for about 40% of total vehicle miles traveled that year (FHWA 2003). In Kansas, the proportion of fatal crashes in rural areas is even higher than the national average. In fact, more than 75% of total fatal crashes in Kansas occurred on rural highways in 2002. Even though these rural highways accounted for 92% of total highway mileage in Kansas, only 53% of vehicle miles traveled occurred on such roadways (FHWA 2003). However, in contrast to fatal crashes, the majority of injury and property-damage-only crashes in Kansas occur in urban areas. These figures indicate the important fact that rural highway crashes result in injuries that are more severe than urban highway crashes and thus, safety of the users of rural highways is one of the crucial issues in improving safety of the overall highway system.

Even though the above figures emphasize the need for a proper agenda to improve highway safety in rural areas, relatively less attention is being paid to the problem. Many factors hamper rural highway safety development efforts. One major challenge is the lack of sufficient funds and resources, particularly crucial due to the huge highway mileage in rural areas. Although, many states are permitted to use their funds for public road safety improvements, the usage of funds is restricted to the development of certain rural highway systems only. On the other hand, local authorities are responsible for the maintenance of most of the rural highways, but they might not be capable of investing large amounts of funding in improving these highways. Moreover, investing large amounts of resources on rural roads might be questionable because of concerns related to cost effectiveness, as these highways account for less traffic volumes compared to urban highways (United States General Accounting Office 2004).

In some cases, crash victims in rural areas become more vulnerable because of delayed response from emergency services. Response time is defined as the time from notification of emergency services until the arrival of EMS personnel at the crash scene. For instance, in the state of Kansas the average emergency service response time for crashes in an urban area is about 13 minutes. The response time for a rural highway crash is about 27 minutes, more than double that of urban areas. This disparity in response times could either be due to the difficulties in reaching the mishap location or the unavailability of emergency services at nearby places. Moreover, in some cases, regardless of whether the road is rural or urban, some states lack the necessary information to make decisions on potential highway safety solutions (United States General Accounting Office 2004).

One way of addressing highway safety issues related to rural highways is by trying to reduce occurrence of crashes by implementing applicable countermeasures. The other way is by trying to reduce the severity of crashes. This approach is particularly important due to the increased severities of crashes in rural areas. However, these two methods can be applied only if the relevant factors contributing to the occurrence and increased severity of crashes are known, making that information important for the highway safety community. Previous similar studies have indicated that these factors could mainly be categorized as driver, environmental, roadway, vehicular or crash related. Although numerous attempts have been made to address the highway safety issues through statistical analysis methods, comparatively fewer studies have analyzed rural highways which result in the majority of fatalities. To the best of our knowledge, no study has dealt with the severity modeling of rural highway crashes.

Accordingly, the objective of this study was to identify the contributing factors likely to affect the severity outcome of rural highway crashes. Identifying these factors would consequently be useful in suggesting countermeasures to reduce the alarming number of high-severity crashes and fatalities in rural areas. This was achieved in this study through statistical modeling of crash severity by using Statistical Analysis Software (SAS 1998).

LITERATURE REVIEW

Various statistical approaches have been utilized to measure the association between various factors and crash severity**.** Shankar et al. (1996) have applied nested logit structure to successfully develop a model to measure the relationship between crash severity and crash prediction factors in rural freeways. The study was based on crash data from a section of Interstate 90 in Washington from 1988 to 1993. The advantage of this method is that the effects of unobserved terms could be avoided as they are cancelled out in the estimation process. The study found that crash severities resulting from run-off-the-road crashes, over-turn crashes, angle crashes, and crashes on curved roadways to be high as compared to others.

In another attempt, Abdel-Aty and Abdelwahab (2004) applied the nested logit structure to investigate the effect of lead vehicle's size on the rear-end crash configuration. Furthermore, the study calibrated different logit nests to estimate the probabilities of rear-end crash configurations as a function of driver age, gender, vehicle type and maneuver, lighting condition, visibility, and speed.

In another study, Ulfarsson and Mannering (2004) applied the nested structure using multivariate multinomial logit models to demonstrate the effect of gender of the vehicle occupant on the severity of injuries suffered in SUV, minivan, pickup, and passenger car crashes.

As a majority of influential factors in highway crashes could be treated more meaningfully as either categorical or dichotomous variables, many researchers have employed categorical data analysis approaches in their studies. A logistic regression modeling approach has been applied by Dissanayake and Lu (2002) to investigate influential factors contributing to older driver injury severity in highway crashes in Florida. Four types of influential factors, which were driver-, environmental-, vehicular-, and highway-related were considered in their attempt to model injury severity. The logistic regression method has also been utilized by many other researchers, such as Farmer and Lund (2002) and Krull et al. (2000), to identify critical factors contributing towards crash severity in different types of highway crashes.

In almost all the crash reporting databases, crash severity is reported in three or more categories, fatal, injury, property damage, etc., enabling the arrangement of severity level from the most severe to the least severe. In other words, severity, which is the response variable in the model, can be considered as an ordinal variable. O'Donnell and Conner (1996) applied this concept to represent the crash severity using both ordered probit and ordered logit structures. The study found that factors such as alcohol involvement, excessive speed, lack of seatbelt usage, head-on collisions and gender of the driver, significantly increase the severity of crashes. In a similar attempt, Khattak et al. (2002) employed an ordered probit modeling approach to investigate the factors contributing to increased injury severities among older drivers. Khattak et al. (2003), Kockelman and Kweon (2002), and Ma and Kockelman (2004) also applied the ordered probit structure in their studies. In an attempt to investigate the contribution of personal and behavioral factors towards injury severity in automobile crashes, Kim et al. (1995) applied log-linear models. In a later study, researchers used these models to study the effect of age, sex, and vehicle type on the driver's responsibility for the crash (Kim et al. 1998).

A review of the studies indicated that many researchers used ordered probit models to examine injury severities in highway crashes (O'Donnell and Connor 1996, Abdel-Aty 2003, and Duncan et al. 1999). However, use of such models to examine severity of rural highway crashes is rare. Thus, this study uses the ordered probit approach to model severity of rural highway crashes in Kansas. In addition, the majority of studies have focused on safety issues related to a particular group of highway users (older or younger drivers, users of a particular vehicle type) or a particular crash type (single vehicle crashes, rear-end crashes) (Abdel-Aty 2003, Ulfarson and Mannering 2004, Dissanayake and Lu 2002, and Farmer and Lund 2002). In such cases, the number of variables that has been taken into account is somewhat limited. Alternatively, the present study tried to identify as many severity-contributing factors as possible, provided they are significant for rural crash severity.

CRASH DATA AND VARIABLE SELECTION

The crash data utilized in this study was extracted from the Kansas Accident Reporting System (KARS) database. The KARS database consists of all data pertaining to highway crashes that occurred on public roadways in Kansas and were reported by police officers. A preliminary analysis was carried out, based on the original data set of 1993 to 2002, to identify the general characteristics of crashes. The results of this analysis showed a rising trend in the number of crashes until 1998 and exhibited a relatively steady pattern later on. The results of the preliminary analysis and the changes made to the coding system of the crash database along with the variations in other characteristics over time were used in selecting a data sample for modeling. For example, all aspects of the transportation system including vehicles, attitudes of drivers, and knowledge of highway users have probably changed over an extended period of time. On the other hand, to have a sufficiently large sample size a longer time span would be needed. Taking all these factors into consideration, data from 1998 to 2002 was selected for statistical modeling. In maintaining the primary focus of this study, data relevant to rural highway crashes was extracted from the KARS database. Each crash record contains driver, vehicular, roadway, and environmental related details along with other crash related details such as crash type, time of occurrence, and emergency response time.

The KARS database records injury severity in five severity levels: fatal, disabling/incapacitating, non-incapacitating, possible injury, and no injury (property-damage-only or PDO). The severity of a crash is identified according to the highest injury sustained by a person involved in the crash. In the statistical modeling approach utilized in this study, severity level pertaining to rural crashes was the dependent variable.

During the data extraction process, crash records involving more than two vehicles, pedestrians, and trains were discarded because the nature of and characteristics related to these crashes are different from other crashes. Additionally, as the subset of data refers to rural crashes, frequencies of such crashes were very small. Some other records were deleted due to missing or incomplete data values which finally resulted in a sample dataset of 93,145 records. Even though one might question the large size of the dataset, all the data was used because it could help avoid any biases resulting from smaller frequencies in certain severity categories. Also, a large sample size would also minimize errors caused by any assumptions made in the modeling process. For example, the normality assumption of the error distribution made in this study could be considered as reliable, because the sample size is large. Part of the selected data sample was randomly separated and used for calibration of the model. Table 1 shows details of some important characteristics related to the crash dataset utilized in the modeling process.

The review of past studies indicated that the majority of studies have focused on safety issues related to specific areas such as a particular group of highway users (older or younger drivers, users of a particular vehicle type) or a particular crash type (single vehicle crashes, rear-end crashes) (Abdel-Aty and Abdelwahab 2004, Ulfarson and Mannering 2004, Dissanayake and Lu 2002, and Farmer and Lund 2002). This study considered almost all rural crashes and tried to identify as many factors contributing to severity as possible, provided they are significant enough to make a difference in the accuracy of the outcome. On the other hand, the quality or the predictability of the statistical model could be expected to increase with an increase in the number of variables.

The candidate factor selection process was based on both the knowledge from previous studies and presumption that a particular factor would be significant to crash severity. Thus, the selected candidate vector of explanatory variables was comprised of many factors and some of these factors may or may not be critical in assessing crash severity. The selected factors and their representation in the model are shown in Table 2. The second column indicates the mean value of the variable estimated by considering the whole data set.

Note that the selection of some of the variables, which were believed to be important, was restricted due to limited availability of data in the database. One such variable was the estimated travel speed of the vehicle at the time of the crash. Many studies have identified travel speed of the vehicle as a significant variable for the severity of the crash (Shanker et al. 1996, Dissanayake and Lu 2002, O'Donnell and Conner 1996, and Khattak et al. 2002). However, in the KARS database the value of this variable was not available for most of the rural crashes, probably because it was difficult for the police officers to make an accurate estimate of the travel speed of vehicles at the time of the crash. Therefore, the posted speed limit at the location of the crash was used in the modeling process instead of travel speed of the vehicle. However, this consideration may lead to over-estimation or under-estimation (generally under-estimation) of the corresponding parameter. Observations based on limited amount of travel speed data revealed that in about 62% of cases travel speed was at or above the posted speed limit. However, in the absence of a better alternative, using posted speed limit could be considered as a satisfactory surrogate measure of actual vehicle speed. Additionally, some other variables such as initial impact point of the vehicle could not be considered in the modeling process due to the lack of detailed information related to those variables.

METHODOLOGY

As shown in Table 2, most of the variables in this study are dichotomous except speed and emergency response time. The dependent or response variable in this study is the crash severity. A variable that can be ranked or ordered, with the difference between two levels being unknown, is an ordinal variable. The response variable in this study, crash severity, can also be ordered as fatal, disabling/incapacitating, non-incapacitating, possible injury, and no injury (PDO). Therefore, crash severity can be considered as an ordinal response variable. A previous publication discussed the applicability of ordered logit and probit models in analyzing this type of data (Long 1997). These ordered choice models are capable of capturing the qualitative difference between two ranked levels, in this case, between two crash severity levels (Khattak et al. 2003).

The difference between the ordered logit and ordered probit structures lies in their distribution assumptions for the unobserved error term. In probit modeling, the error term is assumed to be normally distributed with a mean value of 0 and a variance of 1, whereas the error term in the logit model is assumed to have a logistic distribution with a mean value of 0 and a variance of $p^2/3$, where p= 3.143. Although these methods are based on two different assumptions, they have been found to produce similar results (O'Donnell and Conner 1996).

Table 1: Some Important Characteristics of Crash Data Used for Modeling

Factor	Crash Severity					Total	% *
	Fatal	Incapacitating	Non-incapacitating	Possible	No Injury		
Light Condition							
Day Light	542	1,779	6,181	5,010	29,305	42,817	45.97
Dark	465	1,095	4,199	3,126	41,443	50,328	54.03
Crash Type							
Overturn	239	573	1,749	1,146	2,621	6,328	6.79
Two-vehicle	498	1,097	3,437	3,022	17,882	25,936	27.84
Animal-Vehicle	8	72	609	792	34,961	36,442	39.12
Fixed Object	262	1,132	4,585	3,176	15,284	24,439	26.24
Location							
Intersection	242	703	2,523	2,033	10,140	15,641	16.79
Off the roadway	737	2,057	7,406	5,709	57,614	73,523	78.93
Roadway Related							
Curve / grade	445	1,240	4,215	3,076	23,013	31,989	34.34
Surface wet	127	458	2,014	1,716	13,149	17,464	18.75
Interstate	68	433	1,193	807	7,041	9,542	10.24
Arterial	510	1,253	3,574	2,906	27,516	35,759	38.39
Collector	296	746	3,188	2,434	21,212	27,876	29.93
Local	133	442	2,425	1,989	14,979	19,968	21.44
Speed (mph)**							
< 26	7	32	172	224	2,849	3,284	3.53
26 - 51	86	334	1,873	1,637	12,646	16,576	17.8
51 - 76	914	2,508	8,335	6,275	55,253	73,285	78.68
Emergency Response Time (min)							
<5	141	556	2,266	1,991	17,442	22,396	24.04
5-15	457	1,450	4,889	3,471	23,586	33,853	36.34
15-60	383	829	2,996	2,502	26,041	32,751	35.16
>60	26	39	223	169	3,635	4,092	4.39
Driver Related							
Driver ejected/ trapped	706	951	846	216	72	2791	3
Seat belt not used	720	1,477	3665	2,138	7,536	15,536	16.68
Driver at fault	852	2,332	7,568	5,203	24,470	40,425	43.4
Alcohol /drug Involved	270	515	1,166	559	1,428	3,938	4.23
Total	1,007	2,874	10,380	8,136	70,748	93,145	100
% *	1.08	3.09	11.14	8.73	75.95	100	

** Based on total number of crashes*

*** 1 mph = 1.6 kmph (kilometers per hour)*

Table 2: Explanatory Variables Considered in the Study

Variable	Mean	Description
ALCOHOL	0.04	=1 if alcohol or drug involved, =0 otherwise
ANGLE_CR	0.11	=1 if two vehicles collide at an angle, =0 otherwise
ANM_VEH_CR	0.39	=1 if an animal-vehicle crash, =0 otherwise
ARTERIAL	0.38	=1 if occur on an arterial, =0 otherwise
BLACK_RD_TOP	0.72	=1 if occur on a black road surface, =0 otherwise
COLLECTOR	0.30	=1 if occur on a collector, =0 otherwise
DR_AT_FLT	0.43	=1 if at least one driver is at fault for the crash, =0 otherwise
DR_EJECT	0.03	=1 if at least one driver ejected due to the crash, =0 otherwise
DR_LICENSED	0.97	=1 if driver has a valid license, =0 otherwise
DR_MALE	0.57	=1 if the driver is male, =0 otherwise
DR_NO_STBLT	0.17	=1 if at least one driver not used safety equipments, =0 otherwise
DR_OLD	0.12	=1 if driver age is >55 yrs, =0 otherwise
DR_RESTRICT	0.45	=1 if at least one driver complied with restrictions, =0 otherwise
DR_YOUNG	0.27	=1 if driver age is <25 yrs, =0 otherwise
HDON_CR	0.01	=1 if a head-on crash, =0 otherwise
INTERSTATE	0.10	=1 if occur on an interstate, =0 otherwise
INTR_SECN	0.17	=1 if occur at an intersection, =0 otherwise
LIGHT_CON	0.54	=1 if crash happens in dark or unlit conditions, =0 otherwise
LOCAL	0.21	=1 if occur on a local road, =0 otherwise
ON_RDWAY	0.21	=1 if occur on the roadway, =0 otherwise
PKTIME	0.12	=1 if occur during 6:45 to 9:00 am, =0 otherwise
RD_CUR_GRAD	0.34	=1 if roadway is not straight and level, =0 otherwise
RDCNT_MNT	0.02	=1 if occur at a construction or maintenance zone, =0 otherwise
REAR_END_CR	0.07	=1 if a rear-ended crash, =0 otherwise
RES_TIME	27	Emergency response time in minutes
RES_TIME_BINARY	0. 29	=1 if response time <= 5 minutes, =0 otherwise
ROLLOVER_CR	0.07	=1 if a rollover crash, =0 otherwise
SIDESWIPE_CR	0.04	=1 if a sideswipe crash, =0 otherwise
SNG_VEH_CR	0.33	=1 if a single vehicle crash, =0 otherwise
SPEED	55.12	Speed limit in mph*
TWO_VEH_CR	0.28	=1 if a two-vehicle crash, =0 otherwise
VEH_AT_FLT	0.02	=1 if at least one vehicle is at fault for the crash, =0 otherwise
VEH_AUTMBLE	0.94	=1 if at least one vehicle is an automobile, =0 otherwise
VEH_KS	0.86	=1 if vehicle is registered in Kansas, =0 otherwise
VEH_MNR_STGT	0.72	=1 if vehicle maneuver is straight before crash, =0 otherwise
WEEK_DAY	0.71	=1 if occur on a weekday, =0 otherwise
WET_RD_SURF	0.19	=1 if the road surface wet, =0 otherwise

* *1 mph = 1.6 kmph (kilometers per hour)*

The derivation of the ordered model is based on the measurement model,

(1) $y_i = m \quad if\, \tau_{m-1} \leq y^* < \tau_m$ form = 1 to J

where y* is the injury risk, which is an unobserved continuous variable called latent variable ranging from -∞ to ∞, and is mapped to an observed variable *y*. The τ values are called thresholds or cut-off points and the extreme categories at m = 1 and m = J are defined by open-ended intervals with $\tau_0 = -\infty$ and $\tau_J = \infty$. According to the measurement model, the variable y is perceived to provide incomplete information about an underlying y*.

Then the structural model can be considered as,

(2) $y^* = x_i\beta + \varepsilon_i$

where x_i is a row of a vector of explanatory variables, with an intercept value of 1 in the first column and the i^{th} observation for x_k in the k+1 column. β is a vector of parameters to be estimated and ε_i is the error term, which is assumed to be normally distributed. The KARS database does not contain any information on injury risk (y^*), as it is unobserved. However, the database includes details on the variable *y* observed at different levels of y^*, in which y = 1 if there are no evident injuries, y = 2 if the crash results only in possible injuries, y = 3 when the crash results in non-incapacitating injury, y = 4 if the crash produces incapacitating injury, and y = 5 when the crash is fatal.

Thus, the measurement model can be illustrated as,

$$(3) \quad y_i \begin{cases} 1 \text{ (No injury)} & \text{if } \tau_0 = -\infty \leq y^* < \tau_1 \\ 2 \text{ (Possible)} & \text{if } \tau_1 \leq y^* < \tau_2 \\ 3 \text{ (Non-incapacitating)} & \text{if } \tau_2 \leq y^* < \tau_3 \\ 4 \text{ (Incapacitating)} & \text{if } \tau_3 \leq y^* < \tau_4 \\ 5 \text{ (Fatal)} & \text{if } \tau_4 \leq y^* < \tau_5 = \infty \end{cases}$$

where the threshold values τ_1, τ_2, τ_3 and τ_4 are parameters to be estimated. According to the measurement model, the probability that the i^{th} crash has a severity level of m (m = 1 to 5) is the probability that the injury propensity y^* takes a value between two cut-off points. That is,

(4) $\Pr(y_i = m \mid x_i) = F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta)$

where F(x) is the cumulative distribution function of the unobserved error term ε_I evaluated at a given x under the assumption that ε_I is normally distributed with a mean value of zero and a constant variance, as mentioned earlier. For example, the probability that the victim *I* sustains a fatal injury due to the crash is,

(5) $\Pr(y_i = 5 \mid x_i) = 1 - F(\tau_4 - x_i\beta)$

It should be noted that for these probabilities to be positive the threshold values should satisfy the order, $\tau_1 < \tau_2 < \tau_3 < \tau_4$ (Greene 1997).

The estimation of these model parameters can be carried out through the method of maximum likelihood. The log likelihood, which is the logarithm of the likelihood function, can be written as,

$$(6) \quad \ln L(\beta, \tau \mid y, X) = \sum_{m=1}^{5} \sum_{i=1}^{N} \ln[F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta)]$$

Where N is the total number of observations and β is the vector of parameters from the structural model, in which the first column contains the intercept and τ is the vector of threshold parameters.

The procedure consists of maximizing this equation using numerical methods. To make the model estimable, either one-threshold value, possibly τ_1 or the intercept, is constrained to some arbitrary value, usually zero. The software used in this analysis assumes the intercept $\beta_0=0$ and estimates other parameters. More details on parameter estimation of ordered models using maximum likelihood procedure can be found in textbooks (Long 1997).

The partial change in probability of an i^{th} crash having a severity level of *m,* when a particular influential factor x_k changes, is very useful in interpreting model results. This change is described as marginal effect or a partial change and can be written as,

$$\frac{\partial \Pr(y_i = m \mid x_i)}{\partial x_k} = \frac{\partial F(\tau_m - x_i\beta)}{\partial x_k} - \frac{\partial F(\tau_{m-1} - x_i\beta)}{\partial x_k} \tag{7}$$

In other words, marginal effect is the slope of the probability curve relative to x_k, while holding all other variables constant. The usual practice is to maintain all other variables in their mean values while changing x_k (Long 1997). When there are many dichotomous variables, as in this study, the partial change in x_k becomes meaningless. Therefore, for binary variables analysis is carried out by taking the difference between two probability outcomes (1 and 0) of x_k, while maintaining other variables at their mean values (Long 1997 and Greene 1997).

The R^2 value which is called Generalized Coefficient of Determination is,

$$R^2 = 1 - \left\{\frac{L(0)}{L(\hat{\beta})}\right\}^{\frac{2}{n}} \tag{8}$$

and

$$R^2_{\max} = 1 - \{L(0)\}^{[2/n]} \tag{9}$$

Where *L(0)* is the likelihood of the model and includes only intercept terms. $L(\hat{\beta})$ is the likelihood of the specified model with all the significant factors and *n* is the sample size (Nagelkerke 1991). However, according to Nagelkerke this R^2 value reaches its maximum when it equals a value of 0.75 for models with dichotomous variables, which is the case in this study. This phenomenon contradicts the original definition of the coefficient of determination, identifying the range of R^2 between 0 and 1. Therefore, Nagelkerke (1991) has proposed an adjusted value for R^2, depicted as $\bar{R}^2$, which is defined as,

$$\bar{R}^2 = \frac{R^2}{R^2_{\max}} \tag{10}$$

$\bar{R}^2$ has the minimum and maximum values of 0 and 1 respectively.

MODEL ESTIMATION

When the number of variables is large, as in this study, the amount of time and resources spent for estimating the model is substantial and may lead to some computational burdens. At the same time, the candidate factor selection process was based on prior observations and not on any statistical analysis. Therefore, it is necessary to reduce the number of factors by eliminating non-significant variables or, in other words, by selecting only the factors that are significant on crash severity. O'Donnell and Conner (1996) have used the method of Schwarz Bayesian Information Criteria to accomplish this purpose. This method employs the backward elimination process, starting with all candidate variables and eliminating one at a time by checking the significance of the likelihood ratio.

Instead of applying this method manually, SAS software with a built-in facility for the process of backward selection was used in the analysis (SAS Institute Inc. 1998). In this method, the model starts with all the variables and eliminates the variables with insignificant residual chi-square values one at a time at a given level of confidence (95%). In addition to the backward selection methodology, the software also facilitates stepwise selection, where the model starts without any variables and adds one variable at a time based on the significance of the residual chi-square test. Once a variable is entered into the model it is tested by the backward selection method to make sure that the variable is still significant over other variables present in the model. Both these methods were applied in the model parameter estimation process and they yielded the same results.

At the initial stage of the modeling process, both logit and probit model structures were utilized to model the data with the intention of identifying the better format. The assessment of model results and model fitting information revealed that the probit model structure was more reliable and more capable of predicting crash severity of rural crashes considered in this study. Therefore, the probit model structure was selected for model estimation even though both model structures appeared to be valid.

Initially, the emergency response time was introduced to the model as a continuous variable and the parameter was estimated. However, the estimated parameter for response time was not explaining its effect on crash severity correctly. According to the preliminary analysis of crash data, 95% of all crashes and 97% of injury crashes had an emergency response time of less than one hour. On the other hand, some cases had a response time of more than 20 hours, even though these cases were identified as PDO crashes. This situation may have lead to some unreliable predictions with regard to the parameter. Therefore, it was decided to treat the response time as a categorical variable to obtain a more realistic explanation of its effect on crash severity. Several modeling efforts were carried out using different threshold values of response times and five-minutes was selected as the most appropriate for the data used in this study and the best model was selected as shown in Table 3.

MODEL FITTING INFORMATION

The estimated value of the adjusted R^2 for the final model is 0.38. Thus, the contributing factors in the model are capable of explaining 38% of the variation in crash severity.

Even though there is no generally accepted method for testing the accuracy of ordered multiple-choice models, it is extremely important to check the prediction accuracy of the developed model (O'Donnell and Conner 1996). SAS software produces predicted probabilities for each observation, using the fitted model (SAS Institute Inc. 1998). For example, SAS provides the probability of an observation being fatal, incapacitating, etc., while the predicted overall severity of an observation could be obtained based on the largest individual probability of each severity group. These predicted probabilities were obtained using the fitted model for the subset of the original data sample, which was separated from the original data set. The overall predicted accuracy of the model was found to be 77.9 %. However, the prediction accuracies for different severity categories varied indicating that some severity levels are more difficult to predict than others.

Since there are no other published studies of severity of rural highway crashes, there is no a priori expectation regarding the theoretically expected sign of the explanatory variables.

MODEL RESULTS

Estimated coefficients for the ordered probit model predicting crash severity of rural crashes are shown in Table 3. As the parameter estimation in ordered models assumes a linear relation between the injury risk and explanatory variables (equation 2), interpretation of parameters should be done accordingly. That is, a positive parameter indicates that the relevant variable has an increasing effect on the crash severity, while a negative parameter indicates a decreasing effect on the severity.

Table 3: Maximum Likelihood Estimations of Parameters and Marginal Effects

Factor	Estimated Parameter	Chi-Square Statistic	Marginal Effects				
			Fatal	Incapacitating	Non-Incapacitating	Possible	No Injury
ALCOHOL	0.18	331.21	0.0488	-0.0019	-0.0187	-0.0078	-0.0204
ANGLE_CR	0.438	695.42	0.1069	-0.0087	-0.0411	-0.0161	-0.041
ANM_VEH_CR	-0.244	201.5	-0.0976	-0.0637	-0.0598	-0.0172	0.2383
ARTERIAL	NS	NS	-	-	-	-	-
BLACK_RD_TOP	NS	NS	-	-	-	-	-
COLLECTOR	NS	NS	-	-	-	-	-
DR_AT_FLT	0.151	639.95	0.0359	-0.0004	-0.0136	-0.0059	-0.0159
DR_EJECT	0.813	4877.86	0.2135	-0.0276	-0.0735	-0.0262	-0.0862
DR_LICENSED	-0.058	24.45	-0.0138	0.0003	0.0053	0.0023	0.006
DR_MALE	-0.073	214.97	-0.0174	0.0002	0.0066	0.0029	0.0078
DR_NO_STBLT	0.283	2269.4	0.0684	-0.0037	-0.0263	-0.0107	-0.0277
DR_OLD	0.033	16.09	0.0077	-0.0001	-0.0029	-0.0013	-0.0035
DR_RESTRICT	NS	NS	-	-	-	-	-
DR_YOUNG	NS	NS	-	-	-	-	-
HDON_CR	0.751	1076.58	0.1853	-0.0289	-0.0709	-0.025	-0.0605
INTERSTATE	-0.068	60.54	-0.016	-0.0001	0.006	0.0027	0.0074
INTR_SECN	0.064	26.64	0.0152	-0.0003	-0.0058	-0.0025	-0.0067
LIGHT_CON	NS	NS	-	-	-	-	-
LOCAL	-0.048	47.92	-0.0114	0	0.0043	0.0019	0.0052
ON_RDWAY	-0.07	32.89	-0.0165	0	0.0062	0.0028	0.0076
PKTIME	-0.026	11.29	-0.0061	0	0.0023	0.001	0.0027
RDCNT_MNT	-0.04	6.56	-0.0094	0	0.0035	0.0016	0.0043
RDCUR_GRAD	0.029	33.33	0.0069	-0.0001	-0.0026	-0.0011	-0.0031
REAR_END_CR	0.339	399	0.0824	-0.0059	-0.0317	-0.0126	-0.0323
RES_TIME_BINARY	-0.023	17.06	-0.0054	0	0.002	0.0009	0.0024
ROLLOVER_CR	0.165	399.34	0.0396	-0.0015	-0.0152	-0.0063	-0.0166
SIDESWIPE_CR	0.184	92.37	0.0443	-0.002	-0.017	-0.007	-0.0183
SNG_VEH_CR	0.38	582.08	0.0911	-0.0033	-0.0347	-0.0146	-0.0386
SPEED	0.016	986.86	0.0038	0	-0.0014	-0.0006	-0.0017
TWO_VEH_CR	NS	NS	-	-	-	-	-
VEH_AT_FLT	NS	NS	-	-	-	-	-
VEH_AUTMBLE	NS	NS	-	-	-	-	-
VEH_KS	-0.043	38.95	-0.0103	0.0001	0.0039	0.0017	0.0046
VEH_MNR_STGT	0.064	108.6	0.0151	0	-0.0057	-0.0025	-0.0069
WEEK_DAY	NS	NS	-	-	-	-	-
WET_RD_SURF	-0.123	387.43	-0.029	-0.0003	0.0109	0.0049	0.0135
τ_1	-1.473	332.81	-	-	-	-	-
τ_2	-0.529	43.97	-	-	-	-	-
τ_3	0.519	42.3	-	-	-	-	-
τ_4	0.966	146.55	-	-	-	-	-
R^2	0.308		-	-	-	-	-
Adjusted R^2	0.382		-	-	-	-	-

NS - *Variables are not significant,* - *Not applicable*

The interpretation of marginal effects should be done based on the nature of the corresponding explanatory variable i.e. based on whether the variable is continuous or binary. In the case of continuous variables, a positive marginal effect implies that a unit increase in the explanatory variable from its mean increases the probability of a particular severity level's occurrence by the magnitude of that particular marginal effect while holding other variables at their mean values. For a binary variable, a positive marginal effect implies that the probability of occurrence of a particular severity level increases by the corresponding magnitude of the marginal effect, when the value of the explanatory variable is changed from 0 to 1. However, note that the concept of marginal effects becomes invalid when the value of the variable is far away from its mean.

The following sections consist of discussion of some of the important variables and their effect on rural crash severity.

Driver Related Factors

The positive estimated parameter with a statistically significant chi-square value (significant at 95% confidence level) for the variable 'SPEED' indicates that, with an increase in the posted speed limit, the propensity of suffering a more severe crash also increases. This observation agrees with the findings of several previous studies (even though they do not specifically deal with rural crashes) and is confirmed by positive marginal effects for the fatal severity category as well (O'Donnell and Conner 1996, Dissanayake and Lu 2002, and Khattak et al. 2002). In fact, the probability of incurring a fatal crash increases by 0.004 for a unit increase in speed from its mean value when all the other variables are at their means.

The estimated parameter of the variable relevant to the driver's lack of seatbelt usage (DR_NO_STBLT) is positive. This finding implies that, even if one of the drivers involved in a crash fails to use the seatbelt, the probability of a fatal crash increases by 0.068. The model results indicate that, in the case of driver being thrown out of the vehicle due to the crash (DR_EJECT), the probability of having a high severity crash increases. According to the estimated marginal effects, if the driver is ejected due to the crash, the probability of occurrence of a fatal crash increases by 0.21

In the case of a male driver being involved in a crash, the severity is found to be less, since the variable 'DR_MALE' has a negative estimated parameter. Perhaps this outcome is due to the fact that females are generally less capable of bearing the physical/mental trauma, resulting from the crash, as mentioned by other researchers as well (O'Donnell and Conner 1996).

On the other hand, if at least one of the involved drivers is under the influence of alcohol or drugs, the probability of having a more severe crash is high, as the estimated parameter for variable 'ALCOHOL' is positive. In the KARS database, alcohol involvement has been defined based on whether alcohol presented or alcohol contributed towards the crash based on the judgment of the police officer (Kansas Department of Transportation 2003). However, it should be noted here that in some cases there might not be clear evidence available to make the decision of whether alcohol contributed to the crash or not.

DR_AT_FLT and DR_OLD also have positive estimated parameters. This implies that when at least one driver is at fault for the crash or the involved driver is older than 55 years of age, the probability of having a high severity crash increases. On the other hand, when a driver with a valid driver's license (DR_LICENSED) is involved in a crash the severity can be expected to be low.

Crash Type

Single vehicle crashes (SNG_VEH_CR) has a positive estimated parameter of 0.38, while the variable representing two-vehicle crashes is insignificant. Thus, single vehicle crashes tend to be more severe than two-vehicle crashes. This finding is confirmed by the positive estimated parameter of rollover crashes and the negative estimated parameter for the variable related to crashes occurring on the roadway because the majority of single vehicle crashes are run-off-the-road type crashes. In other words, when the crash occurs off the roadway, there is a greater chance of the crash resulting in higher severity. Animal-vehicle crashes tend to be less severe in nature. According to the KARS database, animal-vehicle crashes account for more than 30% of total rural crashes, most of which are less severe.

Roadway Factors

Roadway geometry (RDCUR_GRAD) results in a positive estimated parameter. This implies that if a crash occurs on a roadway which is not level or straight, the severity of the crash can be expected to be high. According to the model results, the probability of having high-severity crashes on interstate and local roadways is low. On local roads, this may be because there are fewer vehicular interactions with other vehicles. On interstates, better highway attributes and physical features combined with more uniform speeds might lead to this situation. In addition, crashes at intersections are more severe compared to other locations as the variable INTR_SECN has a positive parameter.

Environmental Factors

When a crash occurs on a slippery road surface (under snowy or icy weather conditions) statistical results indicate that the severity of the crash is going to be less, compared to crashes that occur on dry road surfaces (-0.12). Drivers might be paying more attention and be cautious when driving under severe weather conditions and tend to reduce their speeds, which might reduce the possibility of incurring a crash with increased severity. On the other hand, under inclement weather conditions the emergency response time could be a critical factor toward crash severity, because such conditions typically contribute to delayed response from emergency services (Shanker et al. 1996). However, the emergency response time was controlled for in this study, which may indicate the real effects of weather conditions, resulting in more reliable estimations.

Vehicular Factors

When the maneuver of the vehicle before the crash is straight, simply following the road, the probability of having a more severe crash is increased, as the variable 'VEH_MN_STGT' has a positive estimated parameter. The comparison of straight maneuver of the vehicle was made with other types of maneuvers such as right or left turning, U-turning, overtaking, changing lanes, and merging.

According to parameter estimations, when the vehicle (both vehicles in the case of two- vehicle crashes) is registered in the state of Kansas, chance of having a more severe crash is less. This variable was selected with the intention of assessing the effect of driver familiarity with the roads. In other words, out-of-state drivers, unfamiliar with Kansas roads, are more likely to be seriously injured than Kansas drivers.

Emergency Response Time

When the emergency response time is less than five minutes, the possibility of having a crash with more severe injuries is decreased compared to longer response times as the model output shows a negative parameter for this variable. However, note that there was no objective rule in defining this threshold value of five minutes. In fact, even though this cut-off value of five minutes was based on data used in this study, it might be possible to have another threshold value under different conditions. Therefore, a more general interpretation, the longer the emergency response time the higher the probability of having a more severe crash, would be more appropriate. This is confirmed by the marginal probability estimations as the probability of having a fatal crash is decreased by 0.005 when the response time is less than five minutes compared to delayed response times.

Among other factors that affect severity, crashes that occur during peak times (PKTIME) are less severe compared to crashes at other times of the day and crashes in construction or maintenance zones (RDCNT_MNT) are also less severe.

According to the estimated marginal effects, variables related to driver ejection and failure to use seat belts, both have larger marginal effects for fatalities. This implies that having higher seat belt usage would result in significant reduction in severity of rural highway crashes, especially fatal crashes.

CONCLUSIONS

An ordered probit model was developed in this study to identify critical factors contributing to increased crash severity on rural highways. One of the important findings is that the risk of incurring severe injuries is higher when the involved drivers failed to use safety belts at the time of the crash. Since Kansas has a secondary seat belt law, this finding might highlight the need for having a stricter seatbelt law or a primary seatbelt law. It is also noted that there is a higher probability of having a high-severity crash when the driver is ejected from the vehicle due to the crash. It is important to note that when the driver does not wear a seat belt, the probability of ejecting due to the crash is higher. The data used in this analysis were based on police reports and thus the accuracy of the findings is subject to the accuracy of the data used. Particularly in the case of seat belt usage, the accuracy of data is a concern because not everybody may admit to not wearing the seat belt and in many situations the driver might be already out of the vehicle when police officers arrive at the scene.

Factors such as alcohol or drug involvement, posted speed, driver being at fault for the crash, driver being ejected, lack of seatbelt usage, and roadway geometry (not level and straight) appear to augment the severity of rural highway crashes. Crashes that occur on interstate and local roads are less severe. Additionally, single vehicle crashes tend to be highly severe compared to two-vehicle and animal-vehicle crashes. Moreover, with delayed emergency response times the probability of the crash resulting in more severe injuries increases. When a crash occurs under slippery surface conditions of the road or under inclement weather conditions, severity of the crash is found to be less compared to crashes that occur under dry road surface/good weather conditions. Perhaps this outcome is because drivers are more cautious under such adverse conditions and tend to reduce their speeds accordingly.

In general, the study provided some insight to the causes of increased crash severities in rural areas. Findings of the study could be used in suggesting various types of countermeasures to reduce the alarming number of fatalities on rural roadways.

References

Abdel-Aty, M. "Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models." *Journal of Safety Research* 34, (2003): 597-603.

Abdel-Aty, M. and H. Abdelwahab. "Modeling Rear-end Collisions Including the Role of Driver's Visibility and Light Truck Vehicles Using Nested Logit Structure." *Accident Analysis and Prevention* 36 (3), (2004): 447-456.

Dissanayake, S. and J.J. Lu. "Factors Influential in Making an Injury Severity Difference to Older Drivers Involved in Highway Crashes." *Accident Analysis and Prevention* 34 (5), (2002): 609-618.

Duncan, C.S., A. J. Khattak, and F. M. Council. "Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions." *Transportation Research Record* 1635, (1999): 46-54.

Farmer, C. M. and A.K. Lund. "Rollover Risk of Cars and Light Trucks after Accounting for Driver and Environmental Factors." *Accident Analysis and Prevention* 34 (2), (2002): 163-173.

Federal Highway Administration, Office of Highway Policy Information, U.S. Department of Transportation, <http://www.fhwa.dot.gov/policy/ohim/hs02/re.htm> July, 2003.

Greene, W.H. *Econometric Analysis.* Third Edition. Prentice Hall, Upper Saddle River, N.J., 1997.

Kansas Department of Transportation, *Alcohol Involvement in Motor Vehicle Accidents 1993-2003,* http://www.ksdot.org/burTransPlan/prodinfo/2003factsbook/Alcohol.pdf

Khattak, A. J., J.R. Schneider, and F. Targa. "Risk Factors in Large Truck Rollovers and Injury Severity: Analysis of Single-Vehicle Collisions." Transportation Research Record, *Proceedings of 82nd Annual Meeting of the Transportation Research Board* (CD-ROM), Washington D.C., 2003.

Khattak, A.J., M.D. Pawlovich, R.R. Souleyrette, and S.L. Hallmark. "Factors Related to More Severe Older Driver Traffic Crash Injuries." *Journal of Transportation Engineering* 128 (3), (2002): 243-249.

Kim, K., L. Li, R. James, and N. Lawrence. "Drivers at Fault; Influence of Age, Sex, and Vehicle Type." *Journal of Safety Research* 29 (3), (1998): 171-179.

Kim, K., L. Nitz, R. James, and L. Li. "Personal and Behavioral Predictors of Automobile Crash and Injury Severity." *Accident Analysis and Prevention* 34 (5), (1995): 469-481.

Kockelman, K.M. and Y. Kweon. "Driver Injury Severity: An Application of Ordered Probit Models." *Accident Analysis and Prevention* 34 (3), (2002): 313-321.

Krull K. A., A.J. Khattak, and F.M. Council. "Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes." *Transportation Research Record* 1717, (2000): 46-54.

Long, J.S. *Regression Models for Categorical and Limited Dependent Variables.* SAGE Publications, Inc., California, 1997.

Ma, J. and K.M. Kockelman. "Anticipating Injury & Death: Controlling for New Variables on Southern California Highways." *Proceedings of 83rd Annual Meeting of the Transportation Research Board,* Washington, D.C., 2004.

Nagelkerke, N.J.D. "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78, (1991): 691-692.

National Highway Traffic Safety Administration. *Traffic Safety Facts 2002.* U.S. Department of Transportation, 2002.

O'Donnell, C.J. and D.H. Conner. "Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice." *Accident Analysis and Prevention* 28 (6), (1996): 739-753.

SAS Online Doc: Version 8. SAS Institute Inc. Cary, NC, 1998.

Shanker, V., F. Mannering, and W. Barfield. "Statistical Analysis of Accident Severity on Rural Freeways." *Accident Analysis and Prevention* 28 (3), (1996): 391-401.

Ulfarsson, G. F. and F.L. Mannering. "Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents." *Accident Analysis and Prevention* 36 (2), (2004): 135-147.

United States General Accounting Office. *Federal and State Efforts to Address Rural Road Safety Challenges*. Report to Congressional Committees, Publication GAO-04-663 Highway Safety, 2004.

Acknowledgments

This paper is based on research sponsored by the Mack Blackwell Rural Transportation Center at the University of Arkansas. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein.

Sunanda Dissanayke, Ph.D., P.E., *is an assistant professor attached to the Department of Civil Engineering at the Kansas State University. She received her Ph.D. from the University of South Florida and worked there as a research assistant professor before joining the Kansas State University. Her research interests include various aspects of traffic engineering, highway safety, and access management. She has successfully completed several research projects in those areas.*

Indike Ratnayake *is a Ph.D. student in the Department of Civil Engineering at the Kansas State University. He holds a B.S. in civil engineering from the University of Moratuwa, Sri Lanka (2001) and a M.S. in civil engineering (2005) from the Kansas State University.*