Severity of Large-Truck and Combination-Vehicle Accidents in Over-the-Road Service: A Discrete Multivariate Analysis

THIPATAI CHIRACHAVALA, DONALD E. CLEVELAND, and LIDIA P. KOSTYNIUK

ABSTRACT

The severity of large-truck and combination-vehicle accidents was investigated by using 1980 Bureau of Motor Carriers Safety data. The analysis was based on 19,263 accident involvements of such vehicles engaged in over-the-road operation. A two-stage discrete multivariate analysis procedure was used. The first stage involved selecting significant independent variables from 18 candidate variables. The second stage involved modeling accident severity by using two independently estimated logit models, one for the probability of a reported accident being fatal and the other for the probability of a less-severe accident having nonfatal injuries. Differences in the effect of the variables for the four predominant truck types (straight trucks, singles with van, singles with flatbed or tanker, and doubles) led to their separate analyses. The accident severity for each truck type was relatively simple, and most shared common variables and interactions: the main effects of type of collision (single-vehicle or collision with car or commercial vehicle), road class (number of lanes, median separation, and rural or urban setting), and environment (day or night, and road surface wetness condition). The interactions involving road class and environment, and road class and collision type, were usually important. No driver characteristics were found to be significant. Particularly severe accidents were collisions involving passenger cars and doubles, straight trucks, or loaded flatbed or tanker singles on undivided rural roads; collisions involving cars and van singles on undivided rural roads at night; and collisions involving cars and doubles on divided rural roads.

Accidents involving large trucks and combination vehicles are of great concern to the public and to government agencies. The safety issue is more important than ever before because of the 1982 federal legislation that lifted restrictions that long barred the operation of some large combination vehicles in many states. A portion of the findings from a research effort concerning large truck accidents (1) is reported in this paper. The various characteristics that affect the severity of accidents involving at least one large truck or combination vehicle are discussed. A discrete multivariate analysis of a national truck accident data set, including both variable selection and modeling, was carried out.

Why one accident results in a fatality while another does not is the result of a complex interaction that involves the vehicles, the drivers and other occupants, the road, the environment, and chance. Because of such complexity, direct theoretical work has not yet led to sufficient understanding of the problem, and data developed from actual accident experience are analyzed in search of the important determinants of accident severity.

A number of past studies (2-6) compared the accident severity of truck-tractors pulling one trailer (singles) and truck-tractors pulling two trailers (doubles) on the basis of their respective numbers of fatal or injury accidents per million vehicle miles of travel. Hedlund (7), in a study of the effects of road class and vehicle weight on the accident severity of collisions involving combination vehicles and cars, reported that the odds of a fatality varied significantly with the class of roads on which the accidents occurred. Two-lane rural roads were reported to have the highest odds of fatality, followed by four-lane rural roads and urban streets. Hedlund found no independent effect of vehicle weight. No reported studies that simultaneously examine a larger set of possible contributing factors have been found.

DATA

The 1980 Bureau of Motor Carriers Safety (BMCS) data file was used in this research. It contains the records of accidents that involved interstate carriers, subject to the U.S. Department of Transportation Act of 1966 (49 U.S.C. 1655), that resulted in death, injury, or at least $2,000 of property damage. Each accident involvement record includes information on the accident severity, the vehicle (cargo, number of trailers, body style, power unit), driver (age, years with company, hours on duty), environment (light condition, weather), road class (number of lanes, median separation), road surface condition, and accident characteristics. In all, 74 parameters are given for each involvement.

There were 32,245 accident involvements of large trucks and combination vehicles reported to the BMCS in 1980. About 7 percent of these involvements resulted in at least one fatality, 54 percent in at least one nonfatal injury, and the remaining 39 percent in no injury but with property damage exceeding $2,000. Of these, 6,258 accidents involved vehicles engaged in local pickup and delivery service, and the other 25,670 accidents involved vehicles engaged in over-the-road operation. When these accident involvements were broken down by vehicle type, 3,099 involved single-unit trucks (straight trucks), 26,409 involved singles, 1,218 involved doubles, and 1,519 involved tractors only or an unknown vehicle type.

METHODOLOGY

Variable Selection Procedure

The objective of the variable selection was to select a set of significant independent variables from a large number of candidate variables. The variable selection criteria were based on three test statistics--the Pearson chi-square (X_p^2), the total chi-square (Q_T), and the generalized Cochran-Mantel-Haenszel statistic (Q_{CMH}) (8,9). The procedure (8) consisted of the following steps.

1. For each candidate independent variable (V), a Pearson chi-square was computed to test the hypothesis of independence between V and accident severity by using

$$X_p^2 = \Sigma[(\text{Observed} - \text{expected})^2/\text{Expected}] \quad (1)$$

The first variable selected was the one that had the largest X_p^2 per degree of freedom.

2. For each of the remaining variables (U), two-way tables were formed between variable U and accident severity for all levels of the first variable selected. The statistic Q_T was then computed by using

$$Q_T = \sum_{h=1}^{q} \underset{\sim}{G}_h' [\underset{\sim}{Var}(\underset{\sim}{G}_h|H_0)]^{-1} \underset{\sim}{G}_h \quad (2)$$

where $\underset{\sim}{G}_h$ represents a matrix of the differences between the observed and the expected frequencies under H_0 for the hth of q tables, and $\underset{\sim}{G}_h'$ is the transpose matrix of $\underset{\sim}{G}_h$. The null hypothesis for Q_T states that, for each of the separate levels of the covariable set h = 1, 2, ..., q, the response variable is distributed at random with respect to the factor variables. The variable selected was the one with the largest value of Q_T. In this context Q_T reflects both the main effect of a specific variable and its interaction with the previously selected variables. However, after the first few variables were selected, Q_T might lose its asymptotic chi-square property because of the rapidly increasing degrees of freedom and the thinning out of the data in the cells of the contingency table. When this happened, Q_{CMH} could be used instead because it is not affected by the thinning process. It reflects the average partial association effect of a variable, as opposed to Q_T's total contribution, which includes interactions with other variables. Q_{CMH} is expressed as

$$Q_{CMH} = \underset{\sim}{G}' [\underset{\sim}{Var}(\underset{\sim}{G}|H_0)]^{-1} \underset{\sim}{G} \quad (3a)$$

where

$$\underset{\sim}{G} = \sum_{h=1}^{q} \underset{\sim}{G}_h \quad (3b)$$

When using Q_{CMH} as the selection criterion, the variable exhibiting the largest value of Q_{CMH} was chosen.

3. The procedure given in 2 was repeated until the list of candidate variables was exhausted.

Models for Accident Severity

The model form selected for the severity analysis was a pair of logit models for a polytomous severity variable whose categories (fatal, injury, and property damage) have a natural order. Continuous ratios, such as fatal accidents to all nonfatal accidents and nonfatal injury accidents to property-damage accidents, have been shown by Fienberg (10) to possess the asymptotic properties that allow two logit models to be estimated independently while retaining the basic structure of the entire data set. The two logits were (a) that for the probability of fatal accident involvements and (b) that for the probability of nonfatal injury accident involvements. This model form was selected because, although fatal accidents are the ones for which the best understanding is most desired, they account for only a small proportion of total accidents. A single estimated severity model will tend to predict severe accidents far less well than nonsevere accidents.

In a recent study by Gimotty and Chirachavala (11), an estimated logit model for accident severity of passenger-car occupants almost always correctly predicted nonsevere injuries, but it correctly predicted severe injuries less than 50 percent of the time. Another advantage of the continuous ratios is that the independently estimated logits can be relatively more simple than the single severity model. This is desirable because it is easier to interpret a simple model (10). The models for accident severity can be represented by

$P(\text{fatal accident}|\text{an accident}) = f(x_1, x_2, \ldots)$,
and
$P(\text{nonfatal injury accident}|\text{a nonfatal accident}) = f(x_1, x_2, \ldots)$.

In these calculations x_1, x_2, ... are the independent variables.

The logit models for a dependent variable (whose index i = 1, 2, and 3 designates fatal, injury, and property-damage accident involvements, respectively) can be represented by

$$\text{Log}[m_{1jk}/(m_{2jk} + m_{3jk})] = W + W_{2(j)} + W_{3(k)} + W_{23(jk)} \quad (4)$$

$$\text{Log}(m_{2jk}/m_{3jk}) = W + W_{2(j)} + W_{3(k)} + W_{23(jk)} \quad (5)$$

where

- j and k = index levels of variables 2 and 3;
- m_{ijk} = expected cell frequency;
- $m_{1jk}/(m_{2jk} + m_{3jk})$ = ratio of fatal to nonfatal accident involvements; it is called the fatality ratio;
- m_{2jk}/m_{3jk} = ratio of nonfatal injury to property-damage accident involvements; it is called the injury ratio;
- W = overall mean;
- $W_{2(j)}$ = main effect of the jth level of variable 2;
- $W_{3(k)}$ = main effect of the kth level of variable 3; and
- $W_{23(jk)}$ = interaction effect between the jth level of variable 2 and the kth level of variable 3.

The models as represented by Equations 4 and 5 express the probability of a fatal accident involvement given an accident involvement and the probability of a nonfatal injury accident involvement when the involvement is nonfatal.

To ensure that the estimated logit models for the probability of a fatal accident involvement and the probability of an injury accident involvement are integral parts of the same contingency table, the following modeling estimation procedure was adopted. First, a log-linear model for a full contingency

table was estimated to determine the intrinsic associations among all the independent variables. These associations were then included in estimating the logit models for the two probabilities. In estimating the models, the total sample size of the contingency table was assumed fixed.

VARIABLE SELECTION

The 18 independent variables initially considered in the analysis are given in Table 1. They represent accident, vehicle, operational, driver, road, and environmental factors. Accident severity--the dependent variable--was defined to have three levels, as follows:

1. Accident involvements that result in at least a fatality,
2. Accident involvements that result in at least one injury but no fatality, and
3. Accident involvements that result in only property damage that exceeds $2,000.

TABLE 1 Candidate Independent Variables

Variable	Level
Accident: accident type	Single-vehicle, collision with car, collision with commercial vehicle
Vehicle	
Configuration	Straight truck
Trailer body style	Semitrailer, double van, flatbed, tanker, other
Gross vehicle weight	
Load status	Empty, loaded
Operation	
Trip length	Over-the-road, local
Cargo type	General cargo, solid bulk, liquid bulk, metal, chemicals, other
Driver	
Age	
Years of employment	
Hours on duty before accident	
Scheduled time of driving	
Road	
Road class	1-3 lanes or undivided, 4 or more lanes, divided
Road surface condition	Dry, wet, snow
Accident on ramp	Yes, no
Environment	
Rural or urban	
Light condition	Day, night
Time of day	
Weather	Clear, overcast, rain, snow

Ten of the 18 independent variables considered were found to be significant by the variable selection process: accident type, vehicle configuration, trailer body style, gross vehicle weight, cargo weight, trip length, road class, road surface condition, rural or urban environment, and day or night.

The variables that were eliminated (at the 0.05 significant level) were driver age, scheduled time of driving, time of accident, weather, and ramp accident. Driver experience with the company and driver hours on duty were not significant at the 0.01 level and were also eliminated. Cargo type was excluded from further analysis because it was found to be highly correlated with vehicle configuration and trailer body style.

The problems of empty cells and the requirement of reasonable cell counts led to combining some of these 10 variables or the reduction of their number of levels or both. The six final independent variables that were used in the subsequent modeling and their univariate distributions are given in Table 2. The vehicle-type variable consists of four categories: large single-unit truck more than 10,000 lb, truck or tractor with one van-style trailer, truck or tractor with one flatbed- or tanker-style trailer, and tractor pulling two trailers. These four vehicle types will be referred to as straight truck, van single, flatbed or tanker single, and double, respectively. [Not shown in the table is the trip-length variable (local versus over-the-road operations).]

TABLE 2 Truck-Combination Accident Involvements by Variable

Variable	Level	No. of Involvements
Accident type (V2)	Single vehicle	9,891
	With car or motorcycle	10,398
	With commercial vehicle	4,031
Environment (V3)	Day and dry surface	9,389
	Night and dry surface	7,577
	Wet or snowy surface	7,891
Road class (V4)	Rural, 1-3 lanes or undivided	7,925
	Rural, 4 or more lanes divided	8,400
	Urban streets	8,967
Load status (V5)	Empty	7,108
	Loaded	18,888
Vehicle type	Straight truck	774
	Van Semitrailer	13,566
	Flatbed or tanker semitrailer	6,959
	Double	982

Note: Data are from BMCS, 1980.

MODEL ESTIMATION

Although a logit model that has seven categorical variables (six independent variables and one dependent variable) can be estimated, partitioning the data can yield a better understanding of the modeling results. High-order interaction effects in a contingency-table analysis are usually difficult to interpret and typically occur in the situation where there are many variables or the data are unevenly distributed among the cells or both (12).

The first partition of the data was by trip length (local versus over-the-road operations). These two types of operations differ significantly in the environment and operational aspects that may give rise to different severity patterns. Separate analyses were therefore carried out for local and over-the-road operations. Only the findings of over-the-road operation are reported in this paper. The next partition was by vehicle type because, when it was included as one of the variables in the model, a saturated model that exactly duplicated the contingency table was obtained.

The model terms that were found to be significant for the four truck types for the two severity measures are given in Table 3. In most models the main effects of accident type, environment, road class, the interaction between accident type and road class, and the interaction between road class and environment formed the basic common variable effects. Detailed discussion of the significant main effects and interactions is presented for each vehicle type in the following sections. The main effects of the significant variables are highlighted as follows.

The main effect of accident type on both fatality and injury ratios was significant in all eight models (the only such effect). Of the three accident

TABLE 3 Variables and Interactions in Severity Models

FATALITY RATIO MODELS

Vehicle Type	V2	V3	V4	V5	V2xV3	V2xV4	V2xV5	V3xV4	V2xV3xV4
Straight Truck	●	●	●						
Van Semi-Trailer	●	●	●			●		●	
Flat bed/Tanker Semi-Trailer	●	●	●	●	●	●	●	●	●
Double	●		●			●			

INJURY RATIO MODELS

Vehicle Type	V2	V3	V4	V5	V2xV3	V2xV4	V2xV5	V3xV4	V2xV3xV4
Straight Truck	●								
Van Semi-Trailer	●	●	●			●		●	
Flat bed/Tanker Semi-Trailer	●	●	●			●			
Double	●	●	●			●		●	

V2 - Accident Type
V3 - Environment
V4 - Road Class
V5 - Load Status

types, collisions with passenger cars always resulted in the highest fatality and injury ratios. The main effect of environment was significant in six of the eight models. Of the three environmental conditions, dry road surfaces at night usually had the highest fatality and injury ratios. The main effect of road class was significant in seven of the eight models. Of the three road classes, undivided rural roads usually had the highest fatality ratio for all truck types, followed by divided rural roads and urban streets. Hedlund (7) reported a similar finding on the effect of road class. For injury ratios, there was little difference between undivided and divided rural roads. Injury ratios for car-truck collisions on urban streets were as high as those on rural roads.

The interaction between road class and accident type was significant for all six combination-vehicle models. For single-vehicle accidents, there was little difference in the fatality ratios on different road classes, whereas multiple-vehicle collisions had much higher ratios on undivided rural roads. The interaction between road class and environment was significant in four out of six combination-vehicle models. Fatality ratios for wet/snowy pavements were usually lower on divided rural roads than would be expected; dry-surface urban streets during the day also had lower-than-expected fatality ratios.

Severity for Straight-Truck Accidents

There were 640 straight-truck involvements with complete information available on all variables. The models for straight trucks were the simplest of all. The fatality ratios for straight trucks are best explained by the independent main effects of accident type, road class, and environment. The data in Table 4 and Figure 1 give the fitted fatality ratios, which ranged from 0.003 to 0.67. The effect of accident type was such that fatality ratios for single-truck accidents, collisions with a commercial vehicle, and collisions with a car were 1:4:19. The effect of environment was such that fatality ratios for day/dry, wet/snowy, and night/dry were 1:1.5:2.5. The effect of road class was such that the ratios for urban, rural divided, and rural undivided roads were 1:1:5. Note that in this data set two out of five collisions between straight trucks and cars that occur on undivided rural roads at night on dry pavement can be expected to result in a fatality.

The fitted injury-ratio model for straight trucks depends only on the accident type (Table 4). The fitted injury ratio for with-car collisions was more than twice that for either single-vehicle accidents or for with-commercial-vehicle collisions.

Severity for Van Semitrailers

There were 11,748 accident involvements of van singles with complete information on all independent variables. Both the estimated fatality-ratio and injury-ratio models included the interaction between accident type and road class, the interaction between road class and environment, and the main effects of these three variables. The data in Table 5 give the fitted fatality ratios, and Figure 2 shows these effects. Fatality ratios for this vehicle type ranged from 0.01 to 0.26.

The interaction effect between accident type and road class on fatality ratios was such that, for single-vehicle accidents, the ratios for urban, divided rural, and undivided rural roads had a range from 1:1:1 to 0.7:1.6:1, depending on the particular environmental condition. For collisions with cars or with commercial vehicles, the differences in the ratios among these road classes were greater--from 0.4:0.4:1 to 0.3:1:1. The fatality ratios were

TABLE 4 Estimated Fatality and Injury Ratios for Straight Trucks

Road	Environment	Accid. Type	Fatality Ratio	Injury Ratio
Rural Undivided	Dry Day	Single Veh.	0.014	1.13
		With Car	0.259	2.69
		With Comm.	0.054	1.10
	Dry Night	Single Veh.	0.036	
		With Car	0.667	*
		With Comm.	0.139	
	Wet/Snowy	Single veh.	0.023	
		With Car	0.420	*
		With Comm.	0.088	
Rural Divided	Dry Day	Single Veh.	0.003	
		With Car	0.055	*
		With Comm.	0.012	
	Dry Night	Single Veh.	0.008	
		With Car	0.143	*
		With Comm.	0.030	
	Wet/Snowy	Single Veh.	0.005	
		With Car	0.090	*
		With Comm.	0.018	
Urban	Dry Day	Single Veh.	0.003	
		With Car	0.061	*
		With Comm.	0.013	
	Dry Night	Single Veh.	0.008	
		With Car	0.158	*
		With Comm.	0.033	
	Wet/Snowy	Single Veh.	0.005	
		With Car	0.099	*
		With Comm.	0.021	

*Same as entry above

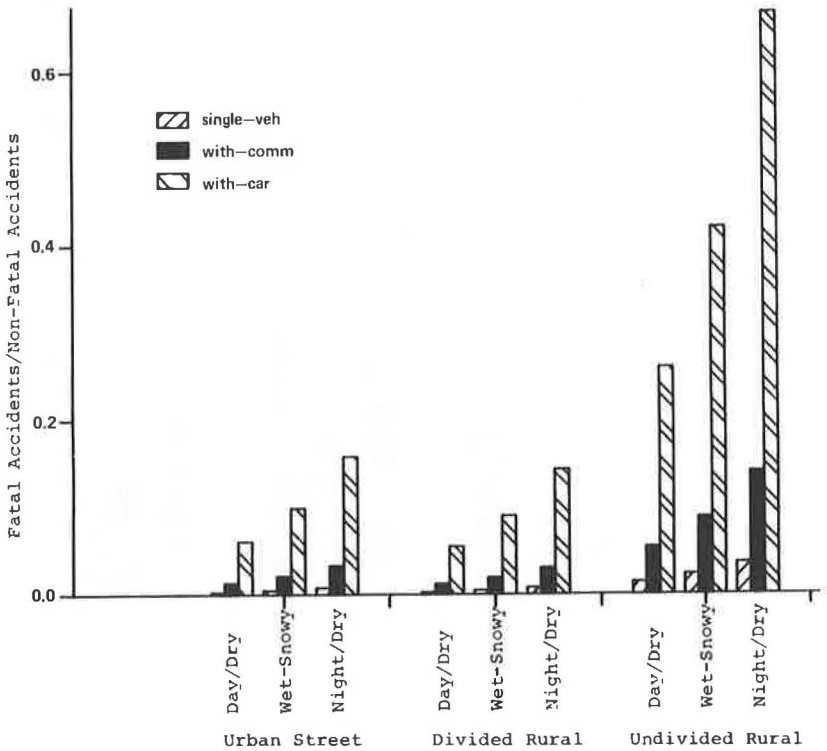

FIGURE 1 Estimated fatality ratios for straight trucks.

TABLE 5 Estimated Fatality and Injury Ratios for Van Semitrailers

Road	Environment	Accid. Type	Fatality Ratio	Injury Ratio
Rural Undivided	Dry Day	Single Veh.	0.023	1.01
		With Car	0.150	2.55
		With Comm.	0.089	1.37
	Dry Night	Single Veh.	0.040	1.31
		With Car	0.261	3.29
		With Comm.	0.155	1.76
	Wet/Snowy	Single veh.	0.022	0.98
		With Car	0.147	2.46
		With Comm.	0.087	1.32
Rural Divided	Dry Day	Single Veh.	0.034	1.14
		With Car	0.102	2.54
		With Comm.	0.079	1.36
	Dry Night	Single Veh.	0.049	1.28
		With Car	0.148	2.86
		With Comm.	0.115	1.54
	Wet/Snowy	Single Veh.	0.021	0.80
		With Car	0.064	1.77
		With Comm.	0.050	0.95
Urban	Dry Day	Single Veh.	0.014	0.50
		With Car	0.035	2.63
		With Comm.	0.022	0.63
	Dry Night	Single Veh.	0.042	0.54
		With Car	0.106	2.86
		With Comm.	0.065	0.69
	Wet/Snowy	Single Veh.	0.019	0.55
		With Car	0.049	2.91
		With Comm.	0.030	0.70

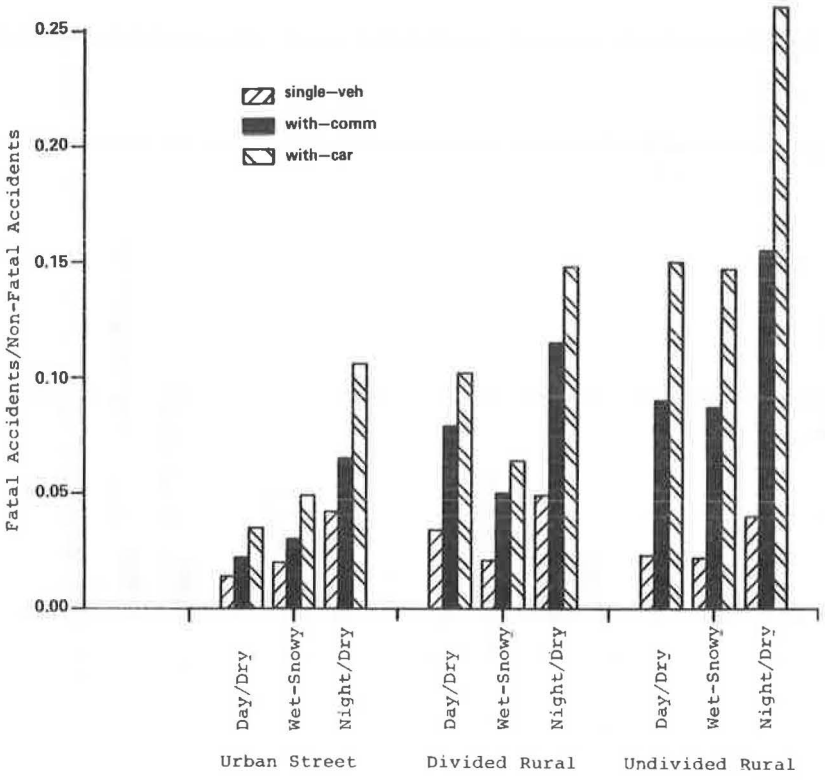

FIGURE 2 Estimated fatality ratios for van singles.

usually the highest for undivided rural roads, followed by divided rural and urban roads.

The interaction effect between environment and road class was such that, for all road classes, there was a much higher fatality ratio for dry pavements at night than for the other two conditions. This is particularly true in two-vehicle collisions. Day accidents on dry pavements had a higher ratio than did the wet/snowy condition for divided rural roads. For undivided rural and urban roads, the difference between these two environmental conditions was small. Of note is a relatively high fatality ratio for collisions with other commercial vehicles on all rural roads at night (0.16 for undivided rural and 0.12 for divided rural roads). These are rare contrasts in which the with-car fatality ratio is not completely dominant for a road class.

The fitted injury ratios for van singles are given in Table 5 and shown in Figure 3. The interaction effect between environment and road class on injury ratios was such that, on all rural roads, there was usually a higher injury ratio at night than during the day on dry pavements; the latter, in turn, was much higher than under wet/snowy condition. For urban roads, all three environmental conditions had similar injury ratios.

The interaction effect between accident type and road class on injury ratios was that, for single-vehicle accidents, the injury ratios on urban, divided rural, and undivided rural roads ranged from 0.5:0.8:1 to 0.5:1:1. For with-car collisions they ranged from 1.1:0.7:1 to 1:1:1. For with-commercial-vehicle collisions they ranged from 0.5:0.7:1 to 0.5:1:1.

Severity for Flatbed or Tanker Semitrailers

There were 6,041 accident involvements of flatbed or tanker semitrailers with complete information on all variables. The fatality ratios are best explained by a two-factor interaction between accident type and load status and a three-factor interaction involving accident type, environment, and road class. The fitted fatality ratios are given in Table 6.

The interaction effect between accident type and load status on fatality ratios was that for single-vehicle and with-commercial-vehicle accidents. Loaded flatbed or tanker singles had twice as high a fatality ratio as did the empty singles, whereas the fatality ratio for with-car collisions did not change much for loaded vehicles. Figure 4 shows this interaction effect on undivided rural roads. Loaded flatbed or tanker singles also had relatively high fatality ratios for with-commercial-vehicle collisions on all rural roads at night on dry pavements.

The three-factor interaction effect among accident type, environment, and road class on fatality ratios was such that dry pavements at night had a higher fatality ratio than did other environmental conditions for all combinations of accident type and road class. The only exception was with-car collisions on undivided rural roads, where there was little difference in fatality ratios for the different environmental conditions.

The injury-ratio model for flatbed or tanker singles included the main effects of accident type, road class, and environment together with the interaction between accident type and road class. The data in Table 6 and in Figure 5 give the fitted injury ratios.

The main effect of environment on injury ratios was that the night/dry condition had higher injury ratios than did the other environmental conditions. The interaction effect between accident type and road class was that single-vehicle accident severity on both types of rural roads was almost 50 percent higher than that found on urban roads. With-car collisions on undivided roads had a similar ratio to

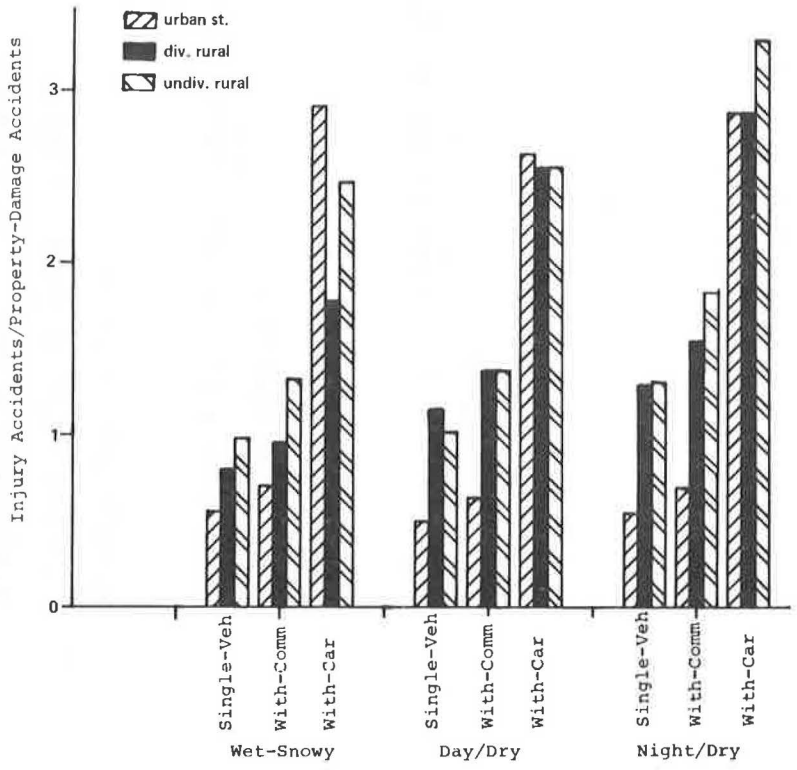

FIGURE 3 Estimated injury ratios for van singles.

TABLE 6 Estimated Fatality and Injury Ratios for Flatbed or Tanker Semitrailers

Road	Environment	Accid. Type	Fatality Ratio		Injury Ratio
			Empty	Loaded	
Rural Undivided	Dry Day	Single Veh.	0.037	0.072	0.78
		With Car	0.153	0.170	2.31
		With Comm.	0.046	0.095	0.96
	Dry Night	Single Veh.	0.047	0.094	1.03
		With Car	0.150	0.166	3.05
		With Comm.	0.069	0.143	1.28
	Wet/Snowy	Single Veh.	0.008	0.016	0.67
		With Car	0.155	0.172	1.98
		With Comm.	0.035	0.071	0.83
Rural Divided	Dry Day	Single Veh.	0.029	0.058	0.74
		With Car	0.086	0.096	2.22
		With Comm.	0.033	0.068	1.23
	Dry Night	Single Veh.	0.039	0.078	0.98
		With Car	0.142	0.158	2.93
		With Comm.	0.069	0.142	1.63
	Wet/Snowy	Single Veh.	0.005	0.011	0.63
		With Car	0.046	0.051	1.90
		With Comm.	0.043	0.087	1.06
Urban	Dry Day	Single Veh.	0.002	0.004	0.51
		With Car	0.054	0.060	2.23
		With Comm.	0.026	0.054	0.51
	Dry Night	Single Veh.	0.023	0.046	0.67
		With Car	0.103	0.114	2.96
		With Comm.	0.028	0.058	0.67
	Wet/Snowy	Single Veh.	0.016	0.033	0.44
		With Car	0.049	0.054	1.92
		With Comm.	0.000	0.000	0.44

those on other roads. The injury ratios for with-commercial-vehicle accidents on urban, divided rural, and undivided rural roads were in a 0.5:1.2:1 proportion.

Severity for Doubles

There were 834 accident involvements of doubles with complete information on all variables. The fatality ratios are best explained by the main effects of accident type and road class, and their interaction. The data in Table 7 and in Figure 6 give the fitted fatality ratios, which ranged in values up to 0.22.

The interaction effect between accident type and road class on fatality ratios was that, for single-vehicle accidents, the ratios on urban, divided rural, and undivided rural roads were about 2:1:1. For with-car collisions they were about 0.2:0.9:1. For with-commercial-vehicle collisions they were 0:0.6:1.

The injury ratio for doubles is best explained by two complex interactions involving road class with accident type and environment. The data in Table 7 and in Figure 7 give the fitted injury ratios. Of interest is that the injury ratios for two-vehicle collisions on wet/snowy urban streets were extremely high (4.8 for with-car and 2.6 for with-commercial-vehicle collisions). Two-vehicle collisions on dry-surfaced divided rural roads at night had considerably higher injury ratios than did those on undivided rural roads. Doubles are the only truck type of which with-car collisions on divided rural roads had a considerably higher injury ratio than did those on undivided rural roads.

SUMMARY AND DISCUSSION OF RESULTS

Of the 18 candidate accident, driver, vehicle, road,

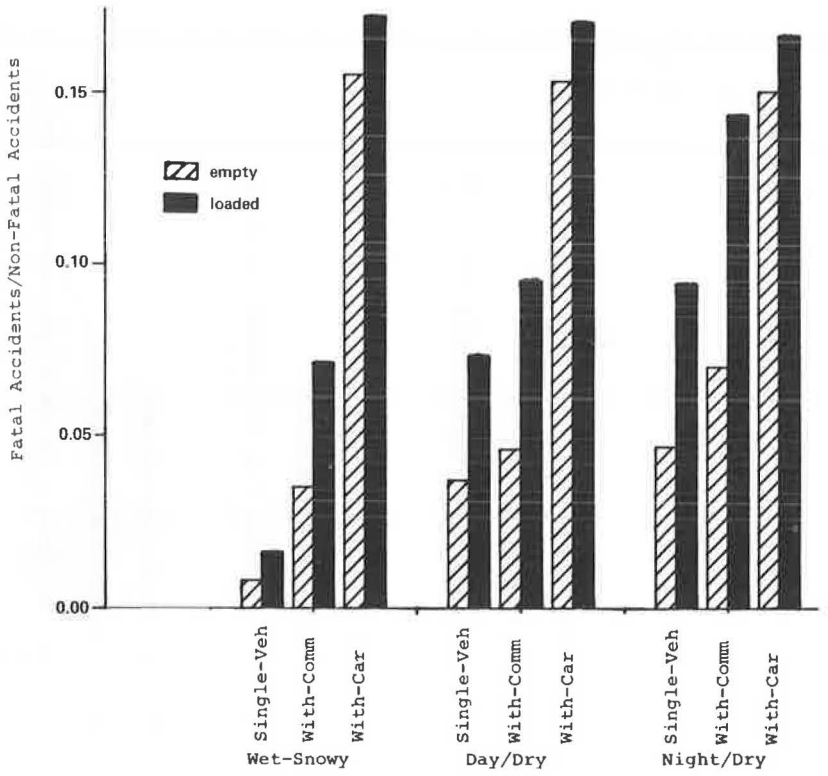

FIGURE 4 Estimated fatality ratios for flatbed or tanker singles.

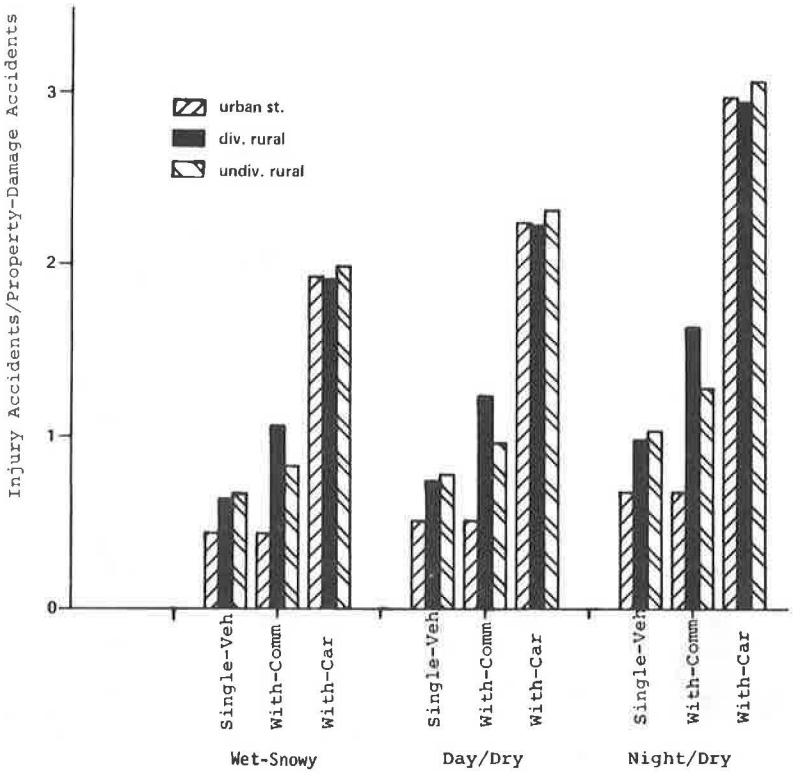

FIGURE 5 Estimated injury ratios for flatbed or tanker singles.

TABLE 7 Estimated Fatality and Injury Ratios for Doubles

Road	Environment	Accid. Type	Fatality Ratio	Injury Ratio
Rural Undivided	Dry Day	Single Veh.	0.031	0.75
		With Car	0.222	2.35
		With Comm.	0.111	1.24
	Dry Night	Single Veh.		0.74
		With Car	*	2.32
		With Comm.		1.23
	Wet/Snowy	Single veh.		0.78
		With Car	*	2.44
		With Comm.		1.30
Rural Divided	Dry Day	Single Veh.	0.042	0.48
		With Car	0.186	1.28
		With Comm.	0.065	0.71
	Dry Night	Single Veh.		1.44
		With Car	*	3.79
		With Comm.		2.10
	Wet/Snowy	Single Veh.		0.98
		With Car	*	2.57
		With Comm.		1.43
Urban	Dry Day	Single Veh.	0.080	0.27
		With Car	0.036	2.01
		With Comm.	0.000	1.06
	Dry Night	Single Veh.		0.28
		With Car	*	2.12
		With Comm.		1.13
	Wet/Snowy	Single Veh.		0.64
		With Car	*	4.77
		With Comm.		2.55

*Same as entry above

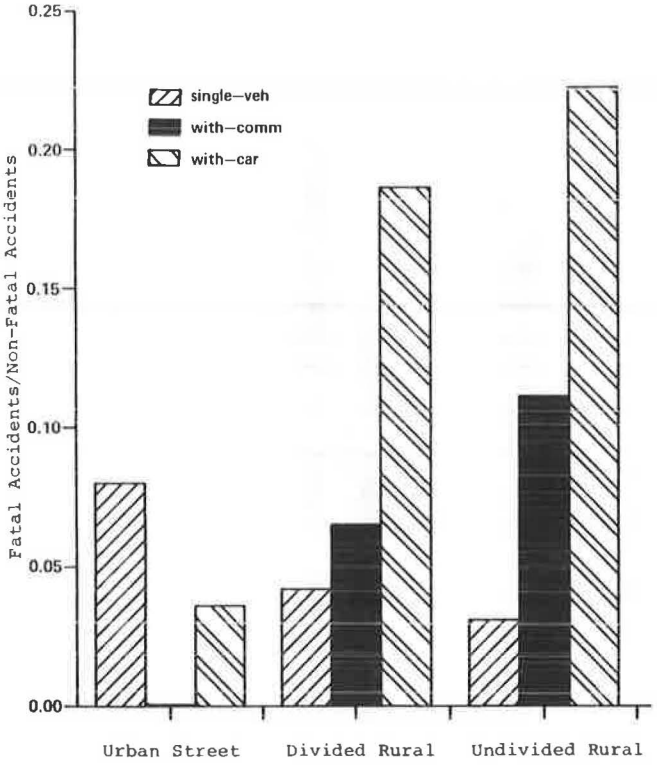

FIGURE 6 Estimated fatality ratios for doubles.

and environment variables considered, those significantly influencing the accident severity of large trucks and combination vehicles engaged in over-the-road operation were vehicle type, accident type, environment, road class, and, to a limited extent, load status. On the other hand, specific driver variables such as age, experience, hours of driving before accident, and scheduled driving time, as well as weather and time-of-day, were not found to be significant. The finding on the effect of road class is generally consistent with that reported by Hedlund (7). The effect of vehicle weight, which had been reported by Hedlund to be nonsignificant, has been found here to be significant for flatbed or tanker singles.

This analysis revealed that relatively simple and similar model structures were effective in capturing the severity differences among the accidents involving most types of large trucks and combination vehicles. The models for the different truck types are easy to interpret, and the differences are large enough to be of practical interest to decision makers.

The models for all combination vehicles, except flatbed or tanker singles, share a basic model form--the main effects of type of collision, road class, environment, and the interactions involving the road class. These effects alone explain most of the variation in accident severity.

The models for straight trucks have no interactions. They include only the main effects of the same three variables: accident type, road class, and environment. The three effects were that fatality ratios for straight trucks were much higher for collisions with cars, for accidents on undivided rural roads, and for accidents on dry pavements at night. The injury ratio for straight trucks was only affected by accident type. With-car collisions had about twice as high an injury ratio as did the other two types of accidents.

Severity models for van singles and doubles shared many common factors--the main effects of accident type, road class, and environment; the interaction between accident type and road class; and the interaction between road class and environment. Fatality ratios of van singles in two-vehicle collisions were higher on undivided rural roads than the

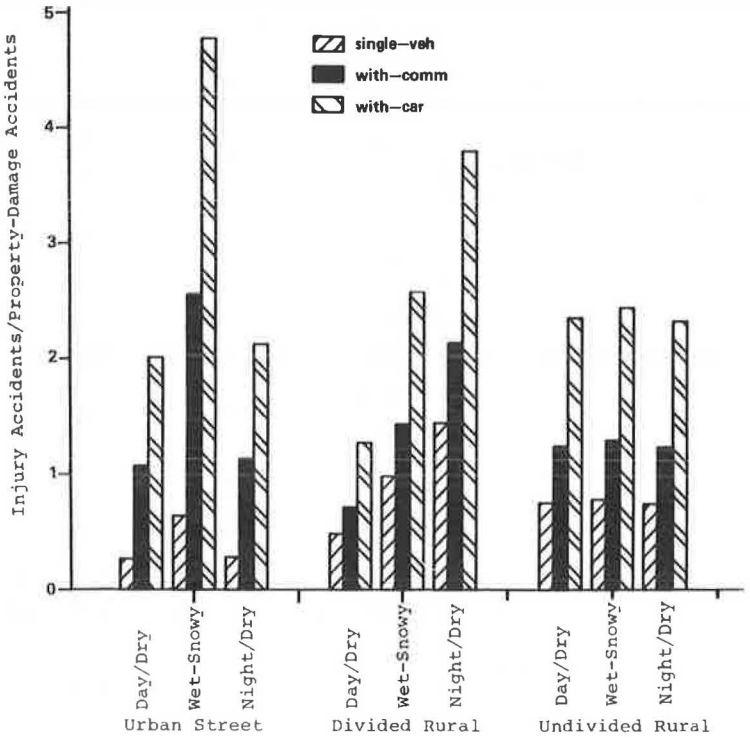

FIGURE 7 Estimated injury ratios for doubles.

other road classes. These collisions also had much higher fatality and injury ratios at night on dry pavements than under the other conditions. Two-vehicle collisions involving at least one double on both types of rural roads had similar fatality ratios. They were about 5 times as high as those found on urban streets. This type of collision on divided rural roads had a higher injury ratio at night on dry pavements than under the other conditions.

The severity models for flatbed or tanker singles were the most complex. In addition to the same three main effects and two interactions for van singles and doubles, the main effect of load status, the interaction between accident type and environment, the interaction between accident type and load status, and the three-way interaction involving accident type, road class, and environment were also significant for the fatality-ratio model for more than 6,000 flatbed or tanker single-accident involvements.

The accident severity for all vehicle types is summarized in Table 8 for the combinations of the other variables. The first symbol in each cell indicates the rating for the fatality ratio, and the second symbol indicates the rating for the injury ratio. The ratings of low, medium, high, and dangerous are based on the following scales of fatality and injury ratios:

1. For fatality ratios: 0 to 0.05 = low (L), 0.06 to 0.15 = medium (M), 0.16 to 0.25 = high (H), and 0.26 to 0.70 = dangerous (D); and
2. For injury ratios: 0 to 1.0 = low (L), 1.1 to 2.0 = medium (M), 2.1 to 3.0 = high (H), and 3.1 to 4.8 = dangerous (D).

The data in Table 8 permit easy identification of the accident characteristics that were most severe:

1. Straight-truck and passenger-car collisions on undivided rural roads under all environmental conditions,
2. Van single and passenger-car collisions on undivided rural roads at night,
3. Double and passenger-car collisions on undivided rural roads under all environmental conditions,
4. Double and passenger-car collisions on divided rural roads under all environmental conditions, and
5. Loaded flatbed or tanker single and passenger-car collisions on undivided rural roads under all environmental conditions.

The vehicle type with the lowest severity record was the empty flatbed or tanker singles. At the other end of the scale were doubles and straight trucks.

For all vehicle types, with-car collisions were the most severe, followed generally by with-commercial-vehicle collisions and single-vehicle accidents. This is because of the vulnerability of occupants of cars when colliding with these large vehicles. This is not surprising because the difference between the weight of a passenger car and the smallest BMCS truck is substantial. With-car collisions on undivided rural roads were usually more severe than those on divided rural roads or urban streets. On undivided rural roads collisions between passenger cars and straight trucks, doubles, and loaded flatbed or tanker singles were always highly severe under all conditions. This is also true for collisions on undivided rural roads at night between passenger cars and van singles. Many of these accidents are head-on. On divided rural roads collisions between passenger cars and doubles were very severe, more so than collisions between cars and other truck types. On urban streets collisions between passenger cars and any truck type resulted in a medium severity level. Prevailing speeds are probably the moderating factor in the severity of these collisions.

Collisions involving commercial vehicles and doubles, loaded flatbed and tanker singles, or van singles on rural roads were usually more severe than those involving commercial vehicles and straight

TABLE 8 Rating of Fatality and Injury Rates for Over-the-Road Operation

			Straight Trucks		Van Semitrailers		Flatbed or Tanker Semitrailers				
							Fatality			Doubles	
Accident Type	Road	Environment	Fatality	Injury	Fatality	Injury	Empty	Loaded	Injury	Fatality	Injury
Single vehicle	Rural divided	Dry day	L	M	L	M	L	M	L	L	L
		Dry night	L	M	L	M	L	M	L	L	M
		Wet/snowy	L	M	L	L	L	L	L	L	L
	Rural undivided	Dry day	L	M	L	L	L	M	L	L	L
		Dry night	L	M	L	M	L	M	L	L	L
		Wet/snowy	L	M	L	L	L	L	L	L	L
	Urban streets	Dry day	L	M	L	L	L	L	L	M	L
		Dry night	L	M	L	L	L	L	L	M	L
		Wet/snowy	L	M	L	L	L	L	L	M	L
With car	Rural divided	Dry day	L	H	M	H	M	M	H	H	M
		Dry night	M	H	M	H	M	M	H	H	D
		Wet/snowy	M	H	M	M	L	L	M	H	H
	Rural undivided	Dry day	D	H	M	H	M	H	H	H	H
		Dry night	D	H	H	D	M	H	H	H	H
		Wet/snowy	D	H	M	H	M	H	M	H	H
	Urban streets	Dry day	L	H	L	H	L	L	H	L	L
		Dry night	M	H	M	H	M	M	H	L	M
		Wet/snowy	M	H	L	H	L	L	M	L	D
With commercial vehicle	Rural divided	Dry day	L	M	M	M	L	M	M	M	L
		Dry night	L	M	M	M	M	M	M	M	H
		Wet/snowy	L	M	L	L	L	M	M	M	M
	Rural undivided	Dry day	L	M	M	M	L	M	L	M	M
		Dry night	M	M	M	M	M	M	M	M	M
		Wet/snowy	M	M	M	M	L	M	L	M	M
	Urban streets	Dry day	L	M	L	L	L	L	L	L	M
		Dry night	L	M	L	L	L	L	L	L	M
		Wet/snowy	L	M	L	L	L	L	L	L	H

Note: L = low, M = medium, H = high, and D = dangerous.

trucks or empty tanker or flatbed singles. However, on urban streets such collisions were more severe when involving doubles or straight trucks than other truck types.

For single-vehicle accidents on rural roads, loaded tanker or flatbed singles had a relatively higher severity level than other truck types. On urban roads doubles had a higher severity level than other truck types.

Note that injury and fatality ratios for accidents on urban streets may conceivably be sensitive to the BMCS reporting process. If the noninjury accidents on urban streets often resulted in property damage less than $2,000, the reported urban-street accidents would have a disproportionately larger fraction of injury and fatality accidents than would the rural-road accidents. This will inflate the injury and fatality ratios for urban streets relative to rural roads. This possible bias, however, will not affect any comparison of severity of the accidents on urban streets made across the other independent variables.

CONCLUSIONS

The need for measures to reduce the severity of car-truck collisions is particularly important for undivided rural roads and urban streets. Such collisions on undivided rural roads usually result in a high incidence of fatality or bodily injuries to car occupants. Although car-truck collisions on urban streets are not usually as severe as those on rural roads, the large number of passenger cars on urban streets can result in a large number of accident casualties.

On undivided rural roads collisions between passenger cars and straight trucks, doubles, and loaded tanker or flatbed singles are highly severe under all conditions, particularly at night. Collisions between passenger cars and van singles are also highly severe at night on undivided rural roads. These point to a need for improved safety measures for nighttime operation of these large vehicles and passenger cars on these roads.

The car-truck collisions on divided rural roads are relatively less serious than those on undivided rural roads, with the exception of collisions between cars and doubles, which have a high severity level. Because doubles are expected to increase in number in the future, special consideration should be given to countermeasures to reduce both the number of such collisions and their severity.

The complex model for flatbed or tanker singles indicates that the severity of the accidents involving this class of vehicles is different from other truck types and warrants further investigation. In particular, attention to measures that will reduce the severity of accidents involving loaded flatbed or tanker singles deserve special consideration because of their higher severity level. Not only is this high severity rate a cause for concern, but because these vehicles are usually engaged in carrying fuels, solid or liquid bulk, logs and poles, metal, and other heavy products, accidents involving these vehicles carry the extra risk of cargo spillage, with possible dangerous effects to the land and communities near the accident site, as well as to other traffic.

ACKNOWLEDGMENT

The authors wish to thank James O'Day of the Transportation Research Institute, University of Michigan, for sharing his insight and knowledge of large-truck accidents with the authors.

Discussion

Olga J. Pendleton*

Chirachavala, Cleveland, and Kostyniuk are to be commended on using a relatively new but powerful and effective statistical technique for analyzing accident severity data. One of the factors that has stifled more extensive use of this method is the availability of computer software for its implementation. It would be informative if the authors would reveal their source of computer software for doing this.

The only points I would like to raise in this discussion involve the variable subset selection process. They are more of an academic nature and are presented as material for future research.

The variable selection procedure described in this study resembles the forward selection procedure in least-squares regression. It is natural to ask, then, if these procedures share the same problems. It is well-known that the forward selection procedure, as applied to regression analysis, may not reveal the best subset because it considers the interrelationship of variables in a sequential manner. That is, if a variable (for example, X_5) is selected as most important according to the the χ^2 test of step 1, it is fixed in the model, and only combinations of that variable and all others are considered as candidate models in the selection process. Suppose a variable that was not significant in step 1 (for example, X_3) would be significant in combination with another variable (for example, X_1), and more significant perhaps than the combination of X_5 and any other variables. The combination of X_3 and X_1 would never be examined as a candidate model by using the current procedure.

The computational problem of examining all possible models is prohibitive both here and in regression. It was circumvented in regression, however, by the discovery of an algorithm that eliminated the necessity of examining all models, but still guaranteed the selection of the best subset based on the criteria of minimum mean squared error (13). Discovery of a similar algorithm for the categorical subset selection process is a fascinating topic for future statistical research.

A second point I would like to raise concerns the treatment of continuous variables in this selection process. Obviously, these variables had to be categorized to accommodate the variable selection criteria. Does this classification affect the variable selection process? For example, if a variable such as driver age was initially classified into equal 20-year age groups and found to be nonsignificant in the variable selection process, would a reclassification of driver age into <20, 20 to 25, 25 to 40, >40 alter the importance of this variable? If so, how should continuous variables be classified? Should a different criteria such as a t-test (or Poisson test) for comparing equality of mean age among severity levels be applied to continuous data? How could that be combined with the categorical selection criteria?

Finally, does the redefinition of categories affect the subset selection process for the discrete variables? In this study classification categories were redefined for the final variables selected to

*Operations Research and Systems Analysis Program, Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843

accommodate zero cells. If the subset selection process were redone by using these new or collapsed categories, would the conclusions of the subset selection process change?

The questions raised in this discussion are difficult to answer but potentially important to the conclusions based on this type of analysis. It is surely beyond the scope of this study; however, this study has provided considerable insight by using a new and powerful statistical tool. In so doing, it has also opened a Pandora's box for statisticians!

Authors' Closure

The discussant is to be thanked for providing a stimulating discussion on the important aspects of the variable selection procedure not explicitly addressed in the paper.

The source of computer software used for the variable selection procedure and the modeling are PARCAT [Landis et al. (9)] and ECTA (Center for Population Studies, University of Michigan). Both are available from the University of Michigan, Ann Arbor, Michigan.

It is true that the variable selection procedure, because of its sequential nature, may not yield the best subset (based on some criteria). However, it is also likely to be true that the selected subset would be close to the best (if it was not the best) because the variables not selected will not significantly contribute to the explanation of the dependent variable in the presence of the selected variables. Besides, the criteria used to evaluate the best statistically often vary with circumstances and individuals' values. Additional test statistics or steps added to those in the existing algorithm that can define and assure the selection of the best subset will certainly be valuable.

When the selection process involves a continuous variable, consideration has to be given to many pertinent questions: how predominant the effect of this variable is relative to the other variables, and whether the effect is truly continuous (i.e., the effect is sensitive to a relatively small change in the values of the variable) or more step-function like. Here the effects of the variables such as driver age, number of years of employment, and scheduled time of driving were more likely to be step-function in nature. Their classifications were aided by past studies, judgment of experts, and examination of the data. In general, the classification adopted for a continuous variable should also be directed at answering (or testing) the substantive questions of interest. It is believed that a t-test and a Poisson test here might not yield conclusive classifications because, at that stage, the effects of other confounding variables were not sufficiently known to permit relatively homogeneous subsets for effective t-tests or Poisson tests. After all, such knowledge was partly what the subsequent modeling task was trying to achieve. In a case where a particular continuous variable is believed to be overwhelmingly important, then an analysis such as that used in Gimotty and Chirachavala (11), which incorporates both continuous and categorical variables, may be considered.

Finally, the redefined categories to accommodate structured zero cells or small cells actually make the modeling results more reliable. This is so because the nature of a chi-square test requires that the expected frequency in each cell be of a reasonable size. The redefinition can conceivably alter the values of the test statistic, the degrees of freedom, and perhaps the P-value of the test if the selection process is redone. But, unless the reclassification is quite extreme or different from the original definition, the significance of a particular variable should not be altered. However, if the researcher is uncertain of the outcome, an analysis such as factor analysis can be used at this step to assist in the reclassification. It is always difficult to generalize the effect of reclassification of variables for all situations without knowing the extent of and the reasons for the reclassification as well as the nature of the variables and the study.

These comments are much more specific to the circumstances that prevailed in this paper. Generalized answers for general situations are not likely to be simple, as the discussant has already pointed out. Statistical tests or models, no matter how sophisticated, are only simplified tools that assist researchers in making inferences on the populations. As such, their applications should go hand in hand with the substantive knowledge in the area of study.

REFERENCES

1. T. Chirachavala. Analysis of Accidents Involving Large Trucks and Combination Vehicles: A Discrete Multivariate Analysis. Ph.D. dissertation. Department of Civil Engineering, University of Michigan, Ann Arbor, 1984.
2. R. Zeiszler. Accident Experience of Double Bottom Trucks in California. California Highway Patrol, Sacramento, April 1973.
3. A Safety Comparison of Doubles Versus Tractor-Semitrailer Operations. Tech. Report. Bureau of Motor Carriers Safety, FHWA, U.S. Department of Transportation, Nov. 1977.
4. S.Y. Chang, M.L. Reiss, and H.W. McGee. Comparison of California Accident Rates for Single and Double Tractor-Trailer Combination Trucks. Final Report. Biotechnology, Inc., Falls Church, Va., March 1978.
5. G.R. Vallette, H.W. McGee, J.H. Sanders, and D.J. Enger. The Effect of Truck Size and Weight on Accident Experience and Traffic Operation--Volume III: Final Report. Biotechnology, Inc., Falls Church, Va., 1982.
6. J.C. Glennon. Trial Record. In The Supreme Court of the United States: Appendix. No. 79-1320. On Appeal from the United States Court of Appeals for the Eighth Circuit, October Term, 1979.
7. J. Hedlund. The Analysis of Large Truck Accidents. Mathematical Analysis Division, National Center for Statistics and Analysis, Research and Development, NHTSA, U.S. Department of Transportation, April 1977.
8. J.E. Higgins and G.G. Koch. Variable Selection and Generalized Chi-Square Analysis of Categorical Data Applied to a Large Cross-Sectional Occupational Health Survey. International Statistical Review, Vol. 45, 1977, pp. 51-62.
9. J.R. Landis, E.R. Heyman, and G.G. Koch. Average Partial Association in Three-Way Contingency Tables: A Review and Discussion of Alternative Tests. International Statistical Review, Vol. 46, 1978, pp. 237-254.
10. S.E. Fienberg. The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge, Mass, 1977.

11. P.A. Gimotty and T. Chirachavala. Injury Models for the Prediction of Severe Accidents. In Proceedings of the 26th Conference of the American Association of Automotive Medicine, 1982, pp. 63-76.
12. Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Mass., 1975.
13. R.R. Hocking and R.N. Leslie. Selection of the Best Subset in Regression Analysis. Technometrics, Vol. 9, 1967, pp. 531-540.

Problem of Identifying Hazardous Locations Using Accident Data

EZRA HAUER and BHAGWANT N. PERSAUD

ABSTRACT

Most agencies with responsibility for extensive road systems use some variant of the rate-and-number method to identify hazardous locations or blackspots. Sites so identified are later examined in detail to diagnose deficiencies and to suggest remedial measures. In this paper the degree to which the rate-and-number method is successful in identifying the unidentified, and what proportion of the sites that are subjected to detailed examination are not deviant at all, is examined. The first part of the paper is devoted to the analysis and development of the mathematical machinery. In the second part the use of the analytical results is illustrated by application to two data sets--one dealing with highway ramps in Ontario and the other with California drivers. The main result of this research is the facility to examine the performance of various identification procedures on the basis of measures of performance that are easy to understand. Such an examination should lead to a realistic assessment of what can be attained when identification for treatment is made on the basis of past accident history.

In most agencies with jurisdiction over extensive road systems it is common practice to try and rectify so-called accident blackspots. Ordinarily a two-stage process is used. In the first stage the past accident history of all sites is reviewed to select a limited number of apparently dangerous locations for further examination. In the second stage the selected sites are studied in more detail, often in the field, in order to devise cost-effective remedial projects for some of the sites.

The two-stage process is required because detailed examination of all sites is impractical. It is hoped that the first stage of the process will act as a sieve. A good sieve is one that allows through all sites that do not require remedial action and retains all sites that do require detailed study. Conversely, an inefficient sieve is one that retains a large number of sites that do not need close scrutiny and allows most blackspots to pass through its holes and thus escape identification. The purpose of this paper is to examine the quality and performance of a commonly used sieve.

Most sieves in current use are a variant of the rate-and-number method. Sites that register an unusually high number of accidents during a specified period of time or an unusually high accident rate (accidents per vehicle kilometer) are selected for inspection. Accidents are often weighted according to their severity. The rationale for the rate-and-number method appears to be left unspecified in the literature. However, a plausible line of reasoning for its raison d'être goes as follows:

> If the accident history of a site is found to deviate from the norm for its class, there surely is some reason for it. If so, a responsible agency and its professionals should examine the cause for this deviation and, if a cost-effective remedy can be found, should remove the cause of substandard performance.

It should be evident that a sieve that screens sites on the basis of number of accidents, accidents per vehicle kilometer, or accident severity is aimed only at establishing deviancy. It is not an indication of how easy or how difficult remedial treatment might be.

Some causes of substandard performance are random and fleeting in nature and essentially unrelated to the physical characteristics of the site. (Consider, for example, a local snow squall or rainstorm that causes several accidents to occur within a few minutes.) Other causes for deviation from the norm are more permanent in nature (sharp curves, polished pavement, narrow bridge, and so forth). These are the causal factors that are subject to remedial action. Accordingly, the object of the exercise is to identify sites for which the deviation from the safety norm is attributable to some permanent properties of the site.