TRANSPORTATION RESEARCH RECORD

No. 1407

Safety and Human Performance
Public Transit

Large-Vehicle Safety Research

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1993

Transportation Research Record 1407
ISSN 0361-1981
ISBN 0-309-05552-0
Price: $20.00

Subscriber Categories
IVB safety and human performance
VI public transit

TRB Publications Staff
Director of Reports and Editorial Services: Nancy A. Ackerman
Associate Editor/Supervisor: Luanne Crayton
Associate Editors: Naomi Kassabian, Alison G. Tobias
Assistant Editors: Susan E. G. Brown, Norman Solomon
Production Coordinator: Sharada Gilkey
Graphics Coordinator: Terri Wayne
Office Manager: Phyllis D. Barber
Senior Production Assistant: Betty L. Hawkins

Printed in the United States of America
National Research Council. Transportation Research Board.

Sponsorship of Transportation Research Record 1407

Group 3–OPERATION, SAFETY, AND MAINTENANCE OF TRANSPORTATION FACILITIES
Chairman: Jerome W. Hall, University of New Mexico

Users and Vehicles Section
Chairman: Conrad L. Dudek, Texas A&M University

Committee on Traffic Records and Accident Analysis
Chairman: Paul P. Jovanis, University of California-Davis
Secretary: Jeffrey F. Paniati, Federal Highway Administration
Sherry S. Borener, David J. Bozak, Thomas E. Bryer, Myung-Soon Chang, Benjamin V. Chatfield, Ted Chira-Chavala, Barbara Hilger DeLucia, Mark Lee Edwards, C. Arthur Geurts, Alan F. Hoskin, Patricia S. Hu, John J. Lawson, Joseph C. Marsh IV, Judson S. Matthias, Carol Lederhaus Popkin, Howard S. Stein, John G. Viner, Phyllis E. Young

Task Force on Statistical Methods in Transportation
Chairman: Olga Pendleton, Texas A&M University System
Secretary: Richard M. Weed, New Jersey Department of Transportation
Michael I. Darter, Gary A. Davis, Michael S. Griffith, Isit Hocherman, Patricia S. Hu, Robert L. Mason, Edward V. Milton, Gary S. Spring, Peter R. Stopher, Robert E. White

Richard Pain, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1992.

Transportation Research Record 1407

Contents

Foreword

Heavy vehicle fatalities may be lower this year than last, but the public eye is still focused on truck and bus crashes. Tracking actual crash experience and putting it in context with that of other highway users is important to identifying needed crash countermeasures, allocating resources to crash reduction, and evaluating progress. The papers by Massie, Campbell, and Blower; Mohamedshah, Paniati, and Hobeika; and Stein increase the ability to do this by providing improved methods, models, and data. Because crashes have multiple causes, understanding the role of the driver and the roadway is imperative to improving the safety of trucks. The papers by Lin, Jovanis, and Yang; and Miaou and Lum address critical issues in these areas.

Modeling the Safety of Truck Driver Service Hours Using Time-Dependent Logistic Regression

TZUOO-DING LIN, PAUL P. JOVANIS, AND CHUN-ZIN YANG

A time-dependent logistic regression model has been formulated to assess the safety of motor carrier operations. The model estimates the probability of having an accident at time interval *t*, subject to surviving (i.e., not having an accident) before that time. Using accident and nonaccident data for 1984 from one national less-than-truckload carrier, nine logistic regression models are estimated that include time-independent effects (i.e., age, experience, multiday driving pattern, and off-duty time before the trip of interest), time main effects (the driving time), and a series of time-related interactions. Driving time has the strongest direct effect on accident risk. The first 4 hr consistently have the lowest accident risk and are indistinguishable from each other. Accident risk increases significantly after the fourth hour, by approximately 65 percent until the seventh hour, and approximately 80 percent and 150 percent in the eighth and ninth hours. The most experienced drivers, those driving more than 10 yr, had the lowest accident risk. All other groups had risks at least 67 percent higher than these safest drivers. There was little difference among the remaining driver groups, although drivers with 1 to 5 yr experience were marginally elevated in risk. Multiday driving patterns had a marginal effect on subsequent accident risk. Daytime driving, particularly in the three days before the day of interest, results in the lowest accident risk. Four driving patterns have an accident risk about 40 to 50 percent higher than Pattern 2: one representing infrequently scheduled drivers; the remaining three involving some type of night driving.

Interstate motor carriers are subject to limitations on the hours that their drivers may be on duty and driving. These include a requirement that a driver be off duty for a minimum of 8 hr after driving for 10 hr or being on duty for 15 hr. There are also cumulative restrictions for on-duty time over several days: 70 hr on duty in 8 days for carriers operating 7 days a week and 60 hr in 7 days for those operating 6 or fewer days a week. These limitations, referred to as the hours-of-service regulations, were initiated in the 1930s. Since then, the U.S. highway system has changed dramatically, as has the nature of the trucking business and the technology of the vehicles. Despite these changes, there have been rather limited attempts to assess the safety implications of the hours of service for contemporary conditions.

Pioneering research was conducted in this area in the 1970s by Harris, Mackie and Miller (*1–3*). Principally using data from accident-involved drivers only, the most enduring finding was a substantial accident risk increase beyond 5 hr of driving. The relationship was derived by comparing the actual number of accidents in each hour of driving with those expected based on the number driving in each hour. This approach accounts for what is called the "survival effect"; that is, a driver who has an accident in the fifth hour successfully completes the first four. Any model of accident risk and driving time must take account of this effect.

Mackie and Miller (*3*) stands out as the most important extant research in the area of multiday driving and accident risk. Interestingly, the most frequent significant declines in performance occurred when the cumulative hour on duty exceeded 70. This indicates that the greatest declines occurred outside the legal driver hour. Recent research (*4,5*) has examined sleeper berth operations and hours of service violations. Others [e.g., Van Der Loop et al. (*6*)] have not adequately included the survival effect in their analysis, compromising their conclusions concerning driving hours.

Harris, Mackie and Miller did not have quantitative statistical modeling methods available to them to study the effects of driver hours of service. Recent biomedical studies have developed the theory and application of such a model using time-dependent logistic regression (*7–10*). A logistic-exponential model (*7*) first suggested that logistic regression could be used for the time-dependent process (e.g., driving time) by dividing time into categories. The model was refined (*8*) to explicitly include time-related interactions and, subsequently, comparisons with the proportional hazard model from survival theory (*9*). A most recent paper developed a method to assess model goodness of fit (*10*).

Earlier motor carrier safety research (*11,12*) has successfully extracted sets of common multiday driving patterns from samples of accident and nonaccident data using cluster analysis. The research reported in this paper builds on earlier studies using survival theory to model motor carrier accident occurrence (*13–15*) by using a larger data set and examining the usefulness of time-related interaction terms with a broader set of models.

OBJECTIVES

Quantitative methods to analyze the effect on accident risk of driver service hours need to be developed. One objective of this paper is to use time-dependent logistic regression to formulate a quantitative model that can include both multiday and consecutive driving time. The second objective is to extensively test the model using data from actual trucking company operations. The models are interpreted with respect to the extant literature and discussed for their policy relevance.

Department of Civil Engineering, University of California, Davis, Davis, Calif. 95616.

LOGISTIC REGRESSION MODEL

A general formulation for the logistic regression model is

$$P(Y_i = 1|X_i) = \frac{\exp[X_i, \beta)]}{1 + \exp[g(X_i, \beta)]} \quad (1)$$

in which Y_i is a response variable representing the occurrence ($Y_i = 1$) or nonoccurrence ($Y_i = 0$) of the event for individual i. X_i is an univariate or multivariate attribute vector for this individual, and $g(X_i, \beta)$ denotes some arbitrary function of X_i and a parameter vector β, which will be estimated. It is implicitly assumed in Equation 1 that the time effect is independent of the covariates. In order to include a time effect, driving time is divided into equal-width intervals. It is not necessary to know the exact time of the accident; accuracy to the level of a specific interval (e.g., 30 min or 1 hr) is sufficient. The time interval in which the accident occurs or the time interval of successful completion of the trip is recorded. A time-dependent logistic regression is therefore formulated (*8,10,16,17*).

Let Y_{ti} be an accident of driver i during the t'th time interval,

$$\begin{aligned} P_{it} &= P(Y_{ti} = 1|Y_{t'i} = 0 \text{ for } t' < t, X_i) \\ &= \frac{\exp[g(X_i, t, \beta)]}{1 + \exp[g(X_i, t, \beta)]} \end{aligned} \quad (2)$$

Equation 2 is the probability of an accident at time interval t, given survival (i.e., no accident) before that time interval. The comparable conditional probability of surviving is defined as

$$Q_{it} = 1 - P_{it} \quad (3)$$

A convenient and simple functional form for $g(X_i, t, \beta)$ is a linear combination of the covariates:

$$g(X_i, t, \beta) = \sum_{j=0}^{r} \beta_j X_{ji} \quad (4)$$

The X_{ji} ($j = 0, \ldots, r$) are the values of the r covariates for the driver i. The full likelihood over the n drivers can be represented by

$$L = \prod_{i=1}^{n} \left(\frac{P_{it_i}}{Q_{it_i}}\right)^{Z_i} \prod_{t'_i \leq t_i} Q_{it'_i} \quad (5)$$

where $Z_i = 1$ for accident driver i, and $Z_i = 0$ otherwise, and t_i represents the number of time intervals for which driver i is exposed to the accident risk.

The addition of the time-dependence parameter (*8*) can be represented as a modification to Equation 4:

$$g(X_i, t, \beta) = \sum_{j=0}^{r} \beta_j X_{ji} + \sum_{k=1}^{T-1} \beta_{r+k} X_{ki^*} \quad (6)$$

X_{ki^*} represents the k'th time interval for driving time. A trip with a length of k time intervals would be represented by a series of indicator variables with $X_{ki^*} = 1$.

This function allows the baseline hazard to vary as a function of time; however, the other covariates are still assumed to be independent of time. Time-dependent effects with other covariates may be added as follows for the m'th variable:

$$X_{ki}^{(m)} = X_{mi} * X_{ki^*} \quad (7)$$

The function will become

$$\begin{aligned} g(X_i, t, \beta) = \sum_{j=0}^{r} \beta_j X_{ji} &+ \sum_{k=1}^{T-1} \beta_{r+k} X_{ki^*} \\ &+ \sum_{s=1}^{T-1} \beta_{r+(T-1)+s} X_{si}^{(m)} \end{aligned} \quad (8)$$

DATA AND VARIABLE DESCRIPTION

Data Collection

The time-dependent logistic regression is conducted using variables that include driver age and experience, the consecutive hours of driving on the trip in question, and the consecutive hours off duty before the last trip. The total number of the observations used for modeling is 1,924 cases, in which 694 cases are accidents. Accidents are deliberately oversampled relative to their actual occurrence in order to more efficiently handle the data.

An accident is defined as any reported event that results in damage to the truck, personal injury, or property damage. Excluded are "alleged" incidents (i.e., those in which someone alleges that they were struck by a truck but no report was filed or verified by the carrier). Because the etiology of these alleged crashes is highly uncertain, it seemed best to ignore these events. Obviously, as in other studies, events that may result in damage but are not reported are not considered. The severity ranges from minor fender-benders to serious injuries, but includes only a few fatalities.

All data are obtained from a national less-than-truckload firm. The company operates "pony express" operations from coast to coast with no sleeper berths. The findings are thus not intended to typify the trucking industry as a whole. As the carrier does take reasonable steps to adhere to U.S. Department of Transportation (U.S. DOT) service hour regulations, the majority of drivers in the study can be considered as not exceeding existing limits.

These data are an expansion of the set used in previous research (*11,12*), which included only the first 6 months of 1984. The analyses presented in this paper use all of the 1984 data set with new cluster analyses and modeling.

Driving Patterns

An important variable in the model is the driving pattern, which includes (a) hours on and off duty over multiple days; (b) the time of day that the on-duty and off-duty hours occurred; and (c) trends of on-duty and off-duty time over several days. A large number of driving patterns are obviously possible over multiple days. In order for this research to succeed, there is a need for a statistical method to identify drivers with similar driving patterns so that the effect of the pattern on risk can be assessed.

Cluster analysis has been successfully used in previous studies to extract common driving patterns (*11,12*). In this research, 10 clusters were selected to describe the driving patterns, an example of which is shown in Figure 1. The proportion of drivers on duty for each 15 min of each of 7 days before the day of interest for one driving pattern is illustrated in this figure. A summary of each driving pattern follows.

Pattern 1: The most frequent on-duty time for this group of drivers occurs from early evening, around 6 p.m., through about 2 a.m. The pattern is highly regular during Days 1, 5, 6, and 7, with more than 80 percent of the drivers on duty at the end of the sixth day and 70 percent during the first, fifth, and seventh days (Figure 1).

Pattern 2: The most frequent on-duty time starts at about 6 a.m. and continues through about 2 p.m. The pattern is highly regular during the last three days, with a peak of 70 percent of the drivers on duty on Days 5, 6, and 7.

Pattern 3: The most frequent on-duty hours are from midnight through about 10 a.m. Hours are regular for the first four days. Driving is rather unlikely during Days 6 and 7.

Pattern 4: The most common on-duty hours begin about 10 a.m. and extend until nearly 6 p.m. Driving becomes very infrequent during Days 5 to 7 but is highly regular during Days 1 to 3.

Pattern 5: The most frequent on-duty time for this group of drivers occurs from evening, around 10 p.m., through morning, about 8 a.m. The pattern is highly regular during Days 1, 2, 6, and 7, and less so during Days 4 and 5.

Pattern 6: The most frequent driving period begins at about 8 p.m., extending until about 6 a.m. Driving is somewhat irregular for Days 1 to 3, but is quite regular over Days 4 to 6.

Pattern 7: The most frequent on-duty times for drivers in this group are from about noon until about 8 p.m. The pattern is quite regular on Days 4 to 7, with nearly 80 percent of the drivers on duty during Days 5 and 6.

Pattern 8: The most frequent on-duty time is from 8 p.m. until 6 a.m. The most frequent on-duty days are 1 through 4.

Pattern 9: The most common on-duty hours begin about 2 p.m. and extend until nearly 10 p.m. Driving becomes very infrequent during Days 5 to 7 but is highly regular during Days 1 to 4.

Pattern 10: This pattern contains drivers who are generally infrequently scheduled, particular during Days 1 to 6.

By inspecting the clusters, several common trends emerge. Pattern 2, 6, and 7 all contain relatively infrequent or irregular driving during the first three or four days but highly regular driving thereafter. Conversely, Patterns 3, 4, 8, and 9 have regular driving during Days 1 to 4 and more irregular driving thereafter. In addition, Patterns 1 and 5 have regular driving during the first two and last two or three days, but infrequent driving during Days 3 to 4.

Data Coding

In order to correctly model the "survival effect," a data duplication method (8,17) is needed because the standard logistic regression model restricts each driver to a trip with only one outcome: an accident or a nonaccident. This procedure is illustrated in Table 1. For a driver with an accident in the third interval, three records will be generated. During the first two records, the values of the response variable would be 0 (nonaccident); whereas for the third record the value of the response variable will be 1. For a driver who successfully completes a trip through the third interval, three records will

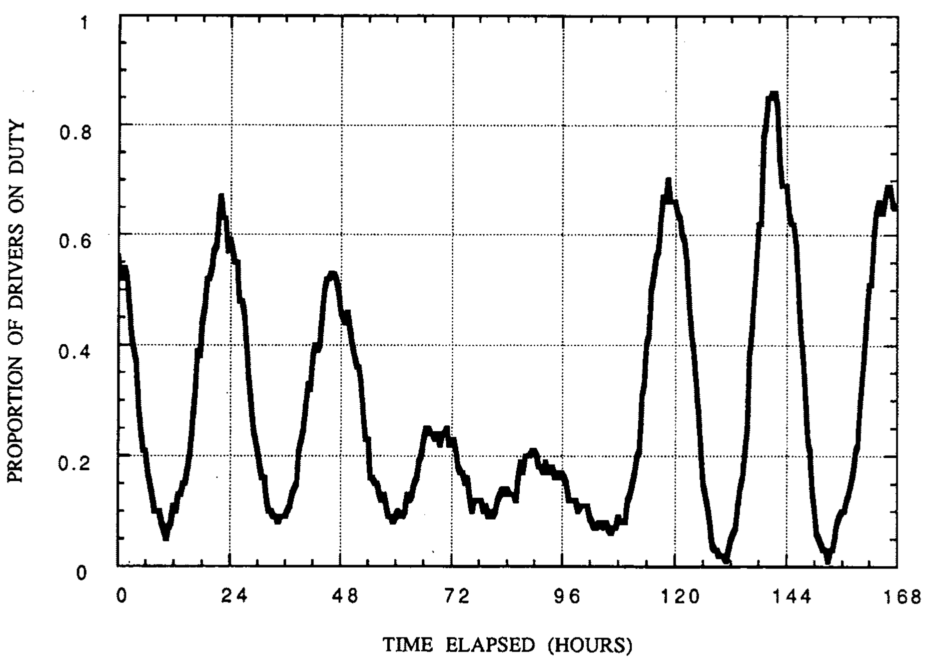


FIGURE 1 Driving Pattern 1.

TABLE 1 Coding Driving Hours and Outcomes for a Survival Effect

CASE 1 : A DRIVER HAS AN ACCIDENT DURING 2-3 HOURS

	COVARIATES	DRIVING HOURS <= 1	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	> 9
NON-ACC	X	1	0	0	0	0	0	0	0	0	0
NON-ACC	X	0	1	0	0	0	0	0	0	0	0
ACC	X	0	0	1	0	0	0	0	0	0	0

CASE 2 : A DRIVER SUCCESSFULLY COMPLETES A 3 HOUR TRIP

	COVARIATES	DRIVING HOURS <= 1	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	8 - 9	> 9
NON-ACC	X	1	0	0	0	0	0	0	0	0	0
NON-ACC	X	0	1	0	0	0	0	0	0	0	0
NON-ACC	X	0	0	1	0	0	0	0	0	0	0
ACC	X	0	0	0	0	0	0	0	0	0	0

also be generated; the values of the response variable for all three records would be 0. The values of the vector of covariates for this individual will be the same in each of the three records. The dummy variables that represent the time effect will be 1 during the time interval to which this record relates, and 0 otherwise. The design variables that represent time-dependent effects with the covariates are coded the same way as those for the time-effect variable, but the values depend on the definition of the type of interaction.

EMPIRICAL RESULTS

Overview of Modeling

An overview of the modeling is contained in Table 2. Models 1 to 3 are developed to separately assess the effect of driving hours, time-independent covariates, and both sets of covariates combined. A series of time-related interactions are estimated in Models 4 to 6. Finally, a large number of additional models are summarized in the discussion of Models 7(a), 7(b), and 8, which attempt to capture the effect of interactions between driving patterns and driving time.

Several tests are conducted to assess the significance of variables and models, including a likelihood ratio test for inclusion or exclusion of a variable as a whole and t-statistics for each category of each variable.

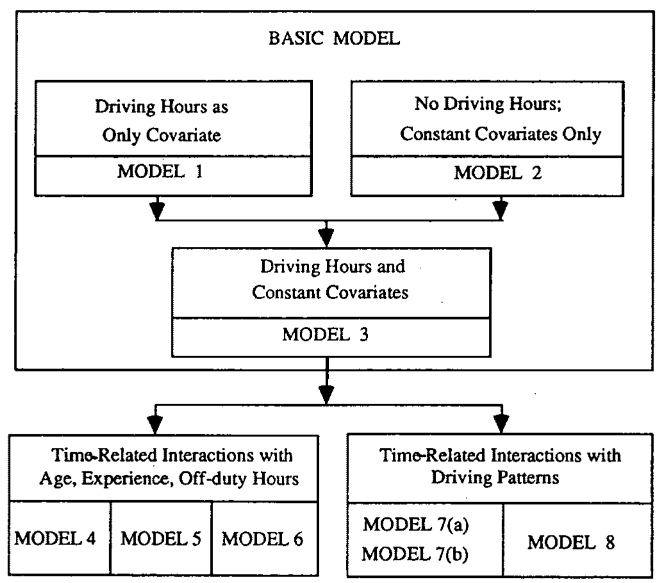
The goodness-of-fit of a model to the data can be qualitatively assessed by plotting model values as a function of driving time against the product limit estimate of the data (10,16). The survival function is denoted as

$$S(t) = \prod_{t' \leq t} Q_{it'} \tag{9}$$

The survival function for the product limit estimator is

$$S(t) = \prod_{t' \leq t} (N_{t'} - D_{t'})/N_{t'} \tag{10}$$

TABLE 2 Modeling Structure



where $N_{t'}$ is the number of drivers at risk at the beginning of the time interval t', and $D_{t'}$ is the number of drivers having an accident during that interval.

Basic Models

Model 1 includes only driving hours, whereas Model 2 includes all other main effects (see Table 3). Model 3 shows the results of combining Model 1 with Model 2. The likelihood ratio test between Model 2 and Model 3 is significant beyond $\alpha = 0.05$, which leads to a rejection of the hypothesis of constant hazard over time. Model 3 is constructed so that there is a constant hazard within each hour and varying hazard between hours. The positive parameter in each covariate represents an increase in the log of the odds ratio or, more simply, an increase in the probability of accident among the drivers in the specific category of the variable compared with the drivers in the corresponding baseline category. The value of the estimated coefficients represent the change in the magnitude of the chance of an accident.

Drivers with more than 10 years experience have the lowest accident risk (baseline category). The risk of other experience levels are all significantly different from the baseline. The highest accident risk occurs when the driving experience is between 1 and 5 years (about 2.2 times higher risk than for the baseline). The estimated risk increase for drivers with less than or equal to 1 year experience and 5 to 10 years are nearly equal (about 1.7 times higher than the baseline category). These results are consistent for all models (1 through 8).

Concerning the multiday driving patterns, Pattern 2, which had the lowest risk, was defined as the baseline driving pattern. The accident risk in Patterns 3, 6, 7, and 10 is significantly different from that in Pattern 2, with a risk about 1.5 times as high. It is interesting that Pattern 10, which contains infrequently scheduled drivers, has an elevated accident risk. Patterns 3 and 6 involve significant night driving, whereas Pattern 7 ends near the end of the peak hours (8 p.m.) and at dusk or night. Drivers who rest less than 9 hr before a trip have a consistent increase in accident risk of about 32 percent. Compared with the baseline of 12 to 24 hours off, this finding is again consistent for all models.

The odds ratio for driving time categories is summarized in Figure 2. The baseline hazard fluctuates from the first hour to the fourth hour with no significant difference, then increases significantly until the last hour. Although the last hour is illustrated in the figure, its estimate is highly uncertain. Examination of the driving hours indicates that nearly 50 percent of the nonaccident trips are completed in the eighth and ninth hours of driving. Because of this high percentage of nonaccident drivers who do not appear in the next time period, they are lost to follow up or have an assumed failure time beyond the completion of their trip. Estimates of the odds ratio in the last driving hour category are thus uncertain and should not be used.

An extensive search of the biostatistics literature produced no comparable empirical problem because most applications involve medical treatments with measurement periods of sev-

TABLE 3 Model Estimates and Statistics

NO	COVARIATES	MODEL 1	MODEL 2	MODEL 3
1	CONSTANT	-3.2780 *	-3.7158 *	-4.0947 *
	AGE			
2	<= 40		0.1381	0.1387
3	40 - 50**			
4	> 50		0.0635	0.0578
	EXPERIENCE (year)			
5	<= 1		0.5174 *	0.5114 *
6	1 - 5		0.7924 *	0.7964 *
7	5 - 10		0.5509 *	0.5677 *
8	> 10**			
	DRIVING PATTERN			
9	1		0.2461	0.2282
10	2**			
11	3		0.3117 *	0.3283 *
12	4		0.2761	0.2984
13	5		0.1430	0.1560
14	6		0.3605 *	0.3773 *
15	7		0.3579 *	0.3677 *
16	8		0.1687	0.1667
17	9		0.2211	0.2324
18	10		0.3269 *	0.3674 *
	OFF-DUTY HOURS			
19	<= 9		0.2593	0.2806 *
20	9 - 12		0.0598	0.0455
21	12 - 24**			
22	> 24		0.1190	0.1141
	DRIVING HOURS			
23	1st HOUR (<1)	0.1404		0.1383
24	2nd HOUR (1 - 2)**			
25	3rd HOUR (2 - 3)	0.1835		0.1894
26	4th HOUR (3 - 4)	0.0040		0.0104
27	5th HOUR (4 - 5)	0.4481 *		0.4630 *
28	6th HOUR (5 - 6)	0.4628 *		0.4812 *
29	7th HOUR (6 - 7)	0.5133 *		0.5396 *
30	8th HOUR (7 - 8)	0.5392 *		0.5788 *
31	9th HOUR (8 - 9)	0.8625 *		0.9128 *
32	10th HOUR (> = 9)	1.8377 *		1.8178 *
	LOG-LIKELIHOOD VALUE	-2698.74121	-2706.63281	-2662.85692
	LIKELIHOOD RATIO TEST (v.s. MODEL 2)			87.55178
	DEGREE OF FREEDOM			9
	CHI-SQUARE (0.95)			16.92

(continued on next page)

TABLE 3 *(continued)*

NO	COVARIATES	MODEL 4	MODEL 5		MODEL 6	
1	CONSTANT	-3.9041 *		-4.4116 *		-4.0703 *
	AGE					
2	<= 40	-0.3308		0.1458		0.1375
3	40 - 50**					
4	> 50	-0.1237		0.0606		0.0591
	EXPERIENCE (year)					
5	<= 1	0.5065 *		1.0965 *		0.5225 *
6	1 - 5	0.7931 *		1.2460 *		0.8031 *
7	5 - 10	0.5656 *		0.8153 *		0.5704 *
8	> 10**					
	DRIVING PATTERN					
9	1	0.2277		0.2280		0.2126
10	2**					
11	3	0.3305 *		0.3291 *		0.3105 *
12	4	0.3097 *		0.3033 *		0.2877
13	5	0.1605		0.1511		0.1465
14	6	0.3790 *		0.3676 *		0.3659 *
15	7	0.3761 *		0.3734 *		0.3577 *
16	8	0.1736		0.1625		0.1486
17	9	0.2364		0.2305		0.2166
18	10	0.3719 *		0.3591 *		0.3576 *
	OFF-DUTY HOURS					
19	<= 9	0.2776 *		0.2865 *		0.9156 *
20	9 - 12	0.0407		0.0416		-0.2289
21	12 - 24**					
22	> 24	0.1138		0.1089		0.0707
	DRIVING HOURS					
23	1st HOUR (<1)	0.0220		0.4170		0.1240
24	2nd HOUR (1 - 2)**					
25	3rd HOUR (2 - 3)	0.0515		0.9438 *		0.0863
26	4th HOUR (3 - 4)	-0.3606		0.1530		-0.0141
27	5th HOUR (4 - 5)	0.1395		0.9340 *		0.6830 *
28	6th HOUR (5 - 6)	0.3801		0.4658		0.5041 *
29	7th HOUR (6 - 7)	0.2593		0.8574 *		0.3852
30	8th HOUR (7 - 8)	0.3238		1.1214 *		0.8078 *
31	9th HOUR (8 - 9)	0.8996 *		0.6079		0.4706
32	10th HOUR (> = 9)	1.2154 *		2.4678 *		1.3066 *
	INTERACTIONS					
33	(2) & (23)	0.1151	(5) & (23)	-0.8759	(19) & (23)	-1.3054 *
34	(2) & (25)	0.1663	(5) & (25)	-1.0819 *	(19) & (25)	-0.6157
35	(2) & (26)	0.9951 *	(5) & (26)	-0.1093	(19) & (26)	-0.2686
36	(2) & (27)	0.5287	(5) & (27)	-0.8954 *	(19) & (27)	-1.0778 *
37	(2) & (28)	0.5300	(5) & (28)	0.0795	(19) & (28)	-1.0512
38	(2) & (29)	0.7721 *	(5) & (29)	-0.4617	(19) & (29)	-0.5412
39	(2) & (30)	0.5611	(5) & (30)	-1.3079 *	(19) & (30)	-1.2632 *
40	(2) & (31)	0.4370	(5) & (31)	-0.2021	(19) & (31)	0.3363
41	(2) & (32)	0.7271	(5) & (32)	-1.1856 *	(19) & (32)	-0.6103
42	(4) & (23)	0.2415	(6) & (23)	-0.3405	(20) & (23)	0.4610
43	(4) & (25)	0.2669	(6) & (25)	-0.9534 *	(20) & (25)	0.5921
44	(4) & (26)	0.1384	(6) & (26)	0.0540	(20) & (26)	0.5259
45	(4) & (27)	0.4679	(6) & (27)	-0.5013	(20) & (27)	-0.1580
46	(4) & (28)	-0.1727	(6) & (28)	-0.6148	(20) & (28)	0.3076
47	(4) & (29)	0.1557	(6) & (29)	-0.4968	(20) & (29)	-0.0599
48	(4) & (30)	0.2677	(6) & (30)	-0.9762 *	(20) & (30)	-0.3378
49	(4) & (31)	-0.3282	(6) & (31)	0.5545	(20) & (31)	0.8501
50	(4) & (32)	1.0083 *	(6) & (32)	-1.1877 *	(20) & (32)	1.1318 *
51			(7) & (23)	-0.0766	(22) & (23)	-0.0290
52			(7) & (25)	-0.9925 *	(22) & (25)	0.0065
53			(7) & (26)	-0.3696	(22) & (26)	-0.3162
54			(7) & (27)	-0.4747	(22) & (27)	-0.2929
55			(7) & (28)	0.2286	(22) & (28)	-0.0376
56			(7) & (29)	-0.2973	(22) & (29)	0.5843
57			(7) & (30)	-0.3305	(22) & (30)	-0.1422
58			(7) & (31)	0.6589	(22) & (31)	0.5522
59			(7) & (32)	-0.4309	(22) & (32)	0.6418
	LOG-LIKELIHOOD VALUE	-2651.669		-2643.57		-2644.93
	LIKELIHOOD RATIO TEST (v.s. MODEL 3)	22.37594		38.57176		35.85446
	DEGREE OF FREEDOM	18		27		27
	CHI-SQUARE (0.90)	25.99		36.74		36.74

(continued on next page)

TABLE 3 (*continued*)

NO	COVARIATES	MODEL 7(a)	MODEL 7(b)		MODEL 8	
1	CONSTANT	-4.3748 *		-4.3785 *		-4.3747 *
	AGE					
2	<= 40	0.1392		0.1412		0.1394
3	40 - 50**					
4	> 50	0.0593		0.0647		0.0587
	EXPERIENCE (year)					
5	<= 1	0.5114 *		0.5114 *		0.5096 *
6	1 - 5	0.7941 *		0.8005 *		0.7971 *
7	5 - 10	0.5697 *		0.5725 *		0.5691 *
8	> 10**					
	DRIVING PATTERN					
9	1	0.5325		0.4934		0.5003
10	2**					
11	3	0.5970		0.5902		0.7352
12	4	0.6012		0.8493		0.6795
13	5	0.4587		0.4190		0.4262
14	6	0.6809		0.6392		0.6478
15	7	0.6717		0.6304		0.6383
16	8	0.4702		0.4313		0.4375
17	9	0.5363		0.4948		0.5026
18	10	0.6706		0.6285		0.6369
	OFF-DUTY HOURS					
19	<= 9	0.2806 *		0.2825 *		0.2812 *
20	9 - 12	0.0467		0.0454		0.0458
21	12 - 24**					
22	> 24	0.1142		0.1128		0.1144
	DRIVING HOURS					
23	1st HOUR (<1)	0.3766		0.4263		0.4115
24	2nd HOUR (1 - 2)**					
25	3rd HOUR (2 - 3)	0.4322		0.5316		0.4631
26	4th HOUR (3 - 4)	0.3204		0.2412		0.2842
27	5th HOUR (4 - 5)	0.7841		0.8044		0.7371
28	6th HOUR (5 - 6)	0.8771 *		0.8484		0.9564 *
29	7th HOUR (6 - 7)	0.8564		0.8211		0.8115
30	8th HOUR (7 - 8)	0.9086 *		0.8270		0.8513
31	9th HOUR (8 - 9)	1.1200 *		1.2568 *		1.1845 *
32	10th HOUR (> = 9)	2.1506 *		1.9919 *		2.0983 *
	INTERACTIONS					
33	(11) & (23)	0.1978	(12) & (23)	-0.4714	(11) & (28)	-2.1560 *
34	(11) & (25)	0.1809	(12) & (25)	-1.0904	(12) & (28)	-1.4327 *
35	(11) & (26)	-0.3607	(12) & (26)	-0.0498	OTHERS	-0.2940
36	(11) & (27)	-0.4801	(12) & (27)	-1.0771		
37	(11) & (28)	-1.9390 *	(12) & (28)	-1.4950 *		
38	(11) & (29)	-0.4217	(12) & (29)	-0.4248		
39	(11) & (30)	-0.5479	(12) & (30)	-0.1420		
40	(11) & (31)	0.3349	(12) & (31)	-1.1514		
41	(11) & (32)	-0.7257	(12) & (32)	0.5947		
42	OTHERS	-0.3306	OTHERS	-0.2851		
	LOG-LIKELIHOOD VALUE	-2654.3252		-2654.6482		-2654.80078
	LIKELIHOOD RATIO TEST (v.s. MODEL 3)	17.0634		16.4175		16.1123
	DEGREE OF FREEDOM	10		10		3
	CHI-SQUARE (0.90)	15.99		15.99		6.25

* t STATISTICS SIGNIFICANT @ α=0.10

** REFERENCED CATEGORY

eral years. Medical subjects typically enter or leave the studies gradually, not with 50 percent departure just before study termination. Truck accident data are likely to have this characteristic. The longer-term solution is to obtain data from firms that can legally operate for longer hours (e.g., in Canada or in California where intrastate driving can occur up to 12 hr consecutively). Of course, the "last" hour is still uncertain but reliable estimates are much more likely for the tenth and eleventh hours.

Comparisons of the survival curves among Model 2, Model 3, and the nonparametric product limit estimator are shown in Figure 3. The survival curve of Model 3 closely follows the trend of the product limit estimator, whereas the non-time-dependent model diverges at mid-range and very high driving times. These findings are consistent with the conclusion that accident risk varies with driving hours: the survival curve for Model 3 bends downward beyond 4 hr, indicating an increase in hazard.

Inclusion of Time-Dependent Interaction Terms

Interaction terms describing the time-dependent effect with covariates are also considered. The purpose is to check the trend of accident risk over time among different categories in each covariate. Models 4 through 6 include interaction terms of driver age, driving experience, and previous off-duty

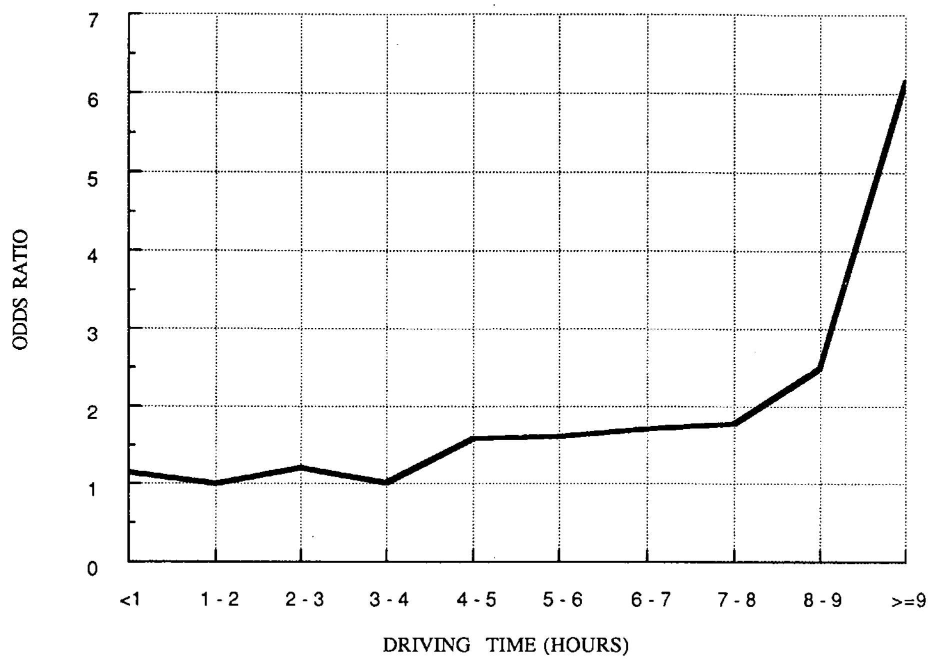


FIGURE 2 Odds ratio of driving hours.

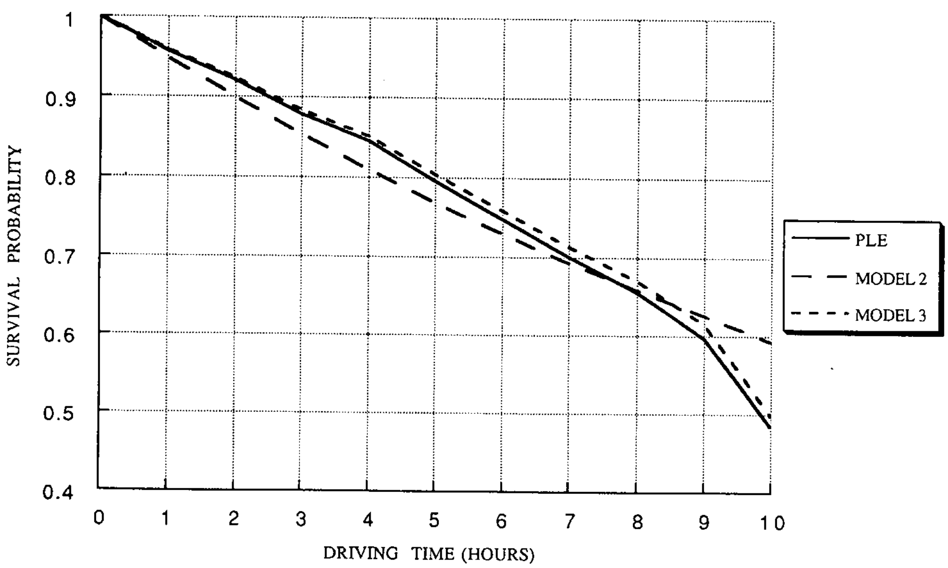


FIGURE 3 Survival curve.

hours, respectively. The likelihood ratio test is applied to test for inclusion or exclusion of the interaction terms. The results show that interactions with driver age and previous off-duty hours as a whole are insignificant beyond $\alpha = 0.10$.

The time-related interactions with driving experience are plotted in Figure 4 as a combined time interaction and main effect. Note the elevated risk for the 1- to 5-year experienced drivers during the first 5 hr of driving. The baseline category, greater than 10 years experience, has consistently lowest risk, particularly for driving hours 5 to 9. This result is consistent with the view that the most experienced drivers are better able to cope with the rigors of long-distance driving, particularly at extended driving times. The improved performance may reflect a learning effect by drivers who may be acquiring the techniques necessary for survival in the traffic stream. There may also be a selection process occurring as only the best drivers are retained over time; the marginal or poor drivers are weeded out by the company as a result of poor driving records or accidents.

It is not practical to include all the interaction terms of driving patterns with the categories of driving hours in one model as 81 additional parameters would have to be estimated. Instead, separate models are developed for time interactions with each driving pattern. The interactions not in-

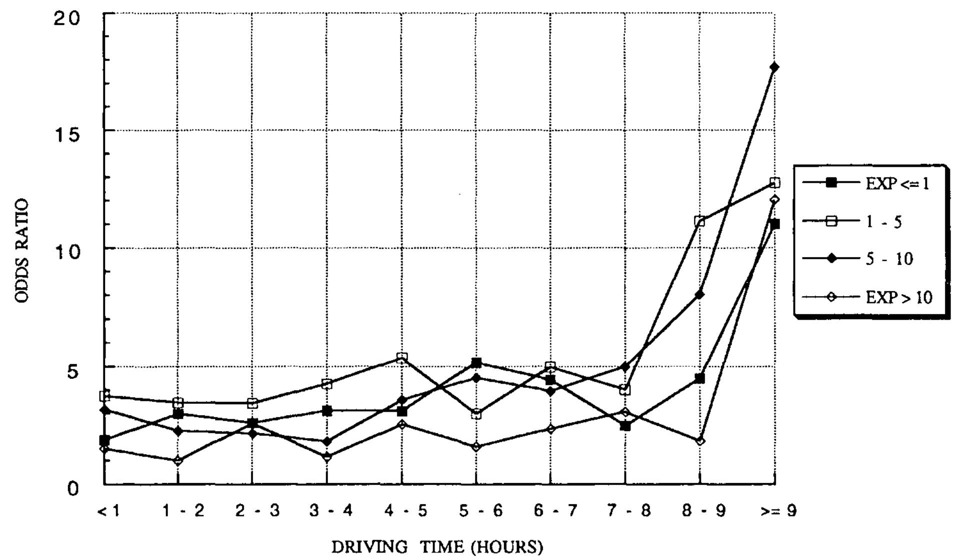


FIGURE 4 Odds ratio of driving experience over time.

cluded in the model are treated as the same effect over time and are combined as one dummy variable. Only the interaction terms of Patterns 3 and 4 with driving hours are significant beyond $\alpha = 0.10$. These are listed in Table 3 as Models 7(a) and 7(b).

The only time-related interaction of significance is a reduction in risk that occurs for Patterns 3 and 4 drivers at a driving time of 5 to 6 hr. It may reflect the benefit of a morning meal break for Pattern 3 and the rest break benefit for Pattern 4 with 5 to 6 hr of driving time. Further modeling of rest break effects is currently under way (18).

Model 8 results from a series of 9 separate models estimated for interactions between driving time and multiday pattern. The interactions with significant t-statistics from each of these models were combined into one model. The interaction parameters that had insignificant t-statistics were excluded step by step; interaction terms with insignificant t-statistics were then treated as the same effect by combining them into one dummy variable. The likelihood ratio test shows that this model is also significantly different from Model 3. This model effectively combines the results of 7(a) and (b). Note that no driving pattern main effects remain significant, indicating the marginal nature of their link to accident risk. Based on the likelihood ratio test and the comparison of the survival curves among these models and nonparametric product limit estimator, Model 8 provides better fit to the data than Models 7(a) or (b).

SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH

A time-dependent logistic regression model has been formulated to assess the safety of motor carrier operations. The model is flexible, allowing the inclusion of time-independent covariates, time main effects, and time-related interactions. The model is used to test the safety implications of current U.S. DOT driving hours of service policies using a data set from a national less than truckload carrier. The model estimates the probability of having an accident at time interval t, subject to surviving (i.e., not having an accident) before that time interval. Covariates tested in the model include consecutive driving time, multiday driving pattern over a 7-day period, driver age and experience, and hours off duty before the trip of interest.

Nine logistic regression models are estimated. Driving time has the strongest direct effect on accident risk. The first 4 hr consistently have the lowest accident risk and are indistinguishable from each other. Accident risk increases significantly after the fourth hour, by approximately 50 percent or more, until the seventh hour. The eighth and ninth hours show a further increase, approximately 80 percent and 130 percent higher than the first 4 hr. These results are generally consistent with those of Harris and Mackie (1).

Driving age and off-duty hours had generally little effect on accident risk except that drivers with 9 or fewer hours off duty before a trip had a 32 percent higher accident risk than drivers with longer off-duty times.

Drivers with more than 10 yr driving experience retain a consistently low accident risk; other categories of driving experience vary a good deal over time. Drivers with 1 to 5 years driving experience, however, have consistently the highest accident risk. Experience with the firm is associated with large changes in risk: a more than doubling of risk for the worst category and a 70 percent increase for the other two.

Multiday driving patterns had a marginal effect on subsequent accident risk. Daytime driving, particularly in the three days before the day of interest (Pattern 2), results in a significantly lower risk on the subsequent day. Four driving patterns have accident risk about 40 to 50 percent higher than Pattern 2; one of these was infrequently scheduled drivers. Two of the remaining multiday patterns involve some type of night driving, whereas the third has the last hours of driving occurring during the peak hours or dusk.

There is general agreement among our findings regarding driving time and those of Harris and Mackie (1), and Mackie and Miller (3). Age appears to play a much less significant role in our accidents whereas experience is much more significant. Multiday driving appears much less significant than in two earlier studies but this may be partially because of the need for even greater precision in driving pattern identification. Subsequent research appears to link the difference in age and experience findings to the inclusion of exposure data. When survival models are estimated without exposure, age is significantly associated with risk; when exposure is added, experience emerges (19).

On the basis of these modeling results, it may be advisable to increase required off-duty time beyond the current 8 hr minimum to something closer to 10 hr. Although the magnitude of the risk increase caused by short off-duty hours is modest, the effect is persistent in all models, attesting to the strength of the association. Although accident risk increases with driving time are clearly substantial, they are particularly disturbing at 8 or 9 hr of driving. Unfortunately this is when the mathematical structure of the model becomes less certain (because of the loss to follow up problem). Our judgment is that this finding will persist when subsequent modeling is conducted, but it weakens our conviction to recommend reducing driver hours regulations.

The effect of multiday driving is much more elusive. Clearly, infrequently scheduled drivers pose a significant risk, providing an incentive for firms to keep drivers busy, albeit legally, and night driving poses some elevated risks. The effect changes somewhat from model to model, occasionally being apparently related more to rest breaks than time of day. There does not appear to be evidence to alter current driver hours policies in this area, although planned ongoing work may be more illuminating.

Further research is needed in areas of model refinement and empirical testing. The addition of roadway-related covariates will greatly aid in separating risk caused by extended driving from risk posed by a change in road design; at least some of the increased risk beyond 5 hr may be explained by terminal access on lower design roads. Work in this area is under way. The driving pattern description may also be refined, to obtain finer resolution of the patterns themselves and to search for patterns that involve shifts in the time of day of driving (e.g., from daylight to early morning or vice versa). The determination of the safest way to change from one driving pattern to another, or the identification of particularly unsafe transitions would be useful information for trucking firms. The effect of rest breaks is the subject of ongoing work (18); development and testing of statistical models for rest effects would be particularly valuable as a guide to trucking operation policies. Modeling and analyzing changes in accident type with driving hours would also be of interest. Analysis of data from truckload, private carrier, or bus operations is also desirable and feasible, given access to appropriate data.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the University of California Transportation Center.

REFERENCES

1. W. Harris and R. R. Mackie. *A Study of the Relationships Among Fatigue, Hours of Service, and Safety of Operations of Truck and Bus Drivers*. Final Report. BMCS RD 71-2, FHWA, U.S. Department of Transportation, 1972.
2. R. Mackie, J. F. O'Hanlon, and M. A. McCanley. *A Study of Heat Noise and Vibration in Relation to Driver Performance and Physiological Status*. Technical Report 1735. Human Factors Research, Inc., Goleta, Calif., 1974.
3. R. Mackie and J. Miller. *Effects of Hours of Service, Regularity of Schedules, and Cargo Loading on Truck and Bus Driver Fatigue*. Report DOT HS-803 799. Human Factors Research, Inc., Goleta, Calif., 1978.
4. R. Hertz. Tractor-Trailer Driver Fatality: The Role of Nonconsecutive Rest in a Sleeper Berth. *Accident Analysis and Prevention*, Vol. 20, No. 6, 1988, pp. 431–439.
5. R. Hertz. Hours of Service Violations Among Tractor-Trailer Drivers. *Accident Analysis and Prevention*, Vol. 23, No. 1, 1991, pp. 29–36.
6. H. Van Der Loop, M. Modde, and I. Veling. Effects from Working-Hour Driving and Rest Periods on Traffic Safety. In *6th World Conference on Transportation Research*, Lyon, France, 1992.
7. M. H. Myers, B. F. Hankey, and N. Mantel. A Logistic-Exponential Model for Use with Response-Time Data Involving Regressor Variables. *Biometrics*, 29, 1973, pp. 257–269.
8. C. C. Brown. On the Use of Indicator Variables for Studying the Time-Dependence of Parameters in a Response-Time Model. *Biometrics*, 31, 1975, pp. 863–872.
9. M. S. Green and M. J. Symons. A Comparison of the Logistic Risk Function and the Proportional Hazards Model in Prospective Epidemiologic Studies. *Journal of Disease*, Vol. 36, No. 10, 1983, pp. 715–724.
10. B. Efron. Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association*, Vol. 83, No. 402, 1988, pp. 414–425.
11. P. P. Jovanis, T. Kaneko, and T. Lin. Exploratory Analysis of Motor Carrier Accident Risk and Daily Driving Patterns. In *Transportation Research Record* 1322, TRB, National Research Council, Washington, D.C., 1991, pp. 34–43.
12. T. Kaneko and P. P. Jovanis. Multiday Driving Patterns and Motor Carrier Accident Risk: A Disaggregate Analysis. *Accident Analysis and Prevention*, Vol. 24, No. 5, 1992, pp. 437–456.
13. H. Chang. *A Disaggregate Survival Model of Motor Carrier Highway Accident Occurrence*. Ph.D. Dissertation, Northwestern University, Evanston, Ill., 1987.
14. P. P. Jovanis and H. Chang. Disaggregate Model of Highway Accident Occurrence Using Survival Theory. *Accident Analysis and Prevention*, Vol. 21, No. 5, 1989, pp. 445–458.
15. H. Chang and P. P. Jovanis. Formulating Accident Occurrence as a Survival Process. *Accident Analysis and Prevention*, Vol. 22, No. 5, 1990, pp. 407–419.
16. R. D. Abbott. Logistic Regression in Survival Analysis. *American Journal of Epidemiology*, Vol. 121, No. 3, 1985, pp. 465–471.
17. D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, Inc., New York, N.Y., 1989.
18. C. Yang, P. P. Jovanis, and T. Lin. *Assessing Motor Carrier Driving Risk Using Cox's Semi-Parametric Model with Multiple Stop Effects*. ITS-Davis Research Report 92-10, University of California, Davis, Calif., 1992.
19. C. Yang, P. P. Jovanis and T. Lin. *Further Tests of Cox's Semi-Parametric Model with Stop Effects*. ITS-Davis Research Report 93-9, University of California, Davis, Calif., 1993 (in preparation).

Publication of this paper sponsored by Task Force on Statistical Methods in Transportation.

Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements

SHAW-PIN MIAOU AND HARRY LUM

Illustrated in this paper are ways in which the Poisson regression model can be used to evaluate the effects of highway geometric design on truck accident involvement rates and estimate and quantify the uncertainties of the expected reductions in truck accident involvements from various improvements in highway geometric design. The data source used in this study was the Highway Safety Information System, a highway safety data base administered by the Federal Highway Administration. Among the five Highway Safety Information states currently available for analysis, Utah was considered to be the state that had the most complete information on highway geometric design and was selected for illustration. Five years of highway geometric, traffic, and truck accident data for rural Interstate highways from 1985 to 1989 were used.

The effects of roadway characteristics on traffic safety are substantial, according to the nation's highway safety performance records (*1*). For example, in 1988 the fatality rates on rural Interstate, other rural Federal-aid primary arterial, and rural non-Federal-aid arterial are, respectively, 9.7, 21.7, and 50.9 fatalities/billion vehicle km (1.56, 3.48, and 8.20 fatalities/100 million vehicle mi). The records also suggest that if all urban and rural travel were at the same fatality rate as the corresponding Interstate rate, then fatalities would be 23,491 instead of 47,093 in 1988, a reduction of over 50 percent (*2*). Potential factors that make vehicle accident rate different from one roadway class to another include the physical nature of the roadway, such as geometric design, roadway markings, and traffic signs, and the type of incurred travel, traffic control, and traffic conditions.

Highway geometric design elements, such as horizontal curvature, vertical grade, lane width, shoulder width, and median, are logical engineering factors that contribute to the differences in vehicle accident rate among roadway classes (*3*). Their effects on vehicle accidents are, however, difficult to quantify because of large confounding influences from the human factor, the environment (including lighting and weather conditions), traffic, and vehicles. Previous studies suggested that roads were rarely the sole factors associated with a traffic accident—only about 2 percent according to Rumar (*4*). It was mainly through the interactions with other factors, especially human and environmental, that roads were associated with traffic accidents.

Ideally, to investigate the effects of highway geometric design on vehicle accidents, roadway, traffic, accident, environmental, road user, vehicle, and exposure data for individual road sections are needed. In practice, many of these factors are qualitative in nature, especially human factors, and are not likely to be available for individual road sections in any accident data base. In addition, detailed vehicle exposure data (e.g., by vehicle type, time of day, weather, and vehicle speed) may not be available for individual road sections. This means that many factors that may have influence on the occurrence of vehicle accidents would not be available for study of the geometric design effects.

In view of this inevitable omitted variable problem, when any geometric design effect is discussed, we have in mind the average observed effect, which includes the collective influence of all the interacting effects. This includes the influence of interacting factors such as the driver's physical condition, driving skill, mood, and knowledge; vehicle speed; weather; and so on. Thus, the geometric design effects are estimated to be conditional on the omitted variables. That is, the effect of the same highway geometric design on vehicle accidents would be different if some of the omitted variables change over time. For example, changes in socioeconomic, legislative, and law-enforcement conditions over the years would change the driver's behavior and, therefore, would change the geometric design effects on vehicle accidents even if nothing is done to the road. For this reason, the analysts should always be careful in interpreting the estimated effects, be conscious of any potential bias, and be cautious in using the effects derived from one area for other areas.

To give another example, consider two hypothetical road sections of the same roadway class, say *C*10 and *C*1, the geometry of which are different only in horizontal curvature: *C*10 is a 10-degree curve (per 30.48-m or 100-ft arc) and *C*1 is a 1-degree curve. The distribution of vehicle speed on *C*10 is expected to be different from that on *C*1, and the average vehicle speed on *C*10 is expected to be less than that on *C*1. Given that the vehicle speed distribution on these two curves is not known, estimated curvature effects on vehicle accidents for *C*10 and *C*1 will be the effects averaged over their respective vehicle speed distribution and, therefore, conditional on their vehicle speed distribution. If the underlying vehicle speed distribution of any curve changes because of speed limit change, for example, then its average curvature effect is likely to change too.

S.-P. Miaou, Center for Transportation Analysis, Energy Division, Oak Ridge National Laboratory, P.O. Box 2008, MS 6366, Building 5500A, Oak Ridge, Tenn. 37831. H. Lum, Office of Safety and Traffic Operations Research and Development, Federal Highway Administration, U.S. Department of Transportation, Turner-Fairbank Highway Research Center, 6300 Georgetown Pike, McLean, Va. 22101-2296.

Many statistical models have been developed to establish the empirical relationships between vehicle accidents and highway geometric design for different roadway classes, vehicle configurations, and accident severity types (3). However, most of these models were developed on the basis of the conventional multiple linear regression approach, and have been shown to be lacking the distributional property to adequately describe discrete, nonnegative, and typically sporadic vehicle accident events on the road (5–7). These unsatisfactory properties of the linear regression models have led to the investigation of the Poisson regression and negative binomial regression models in recent studies (6–9). In general, most of these studies found the Poisson regression model to be appropriate for studying the relationships among vehicle accidents and the contributing factors under their study. In addition, despite the limitations in existing highway geometric data, some encouraging relationships have been developed for horizontal curvature, vertical grade, and shoulder width using the Poisson regression model.

The objective of this paper is to illustrate how the Poisson regression model can be used to evaluate the effects of highway geometric design elements on truck accident involvement rates. Also, described in this paper is the way in which the model can be used to estimate and quantify the uncertainties of the expected reductions in truck accident involvements from various improvements in highway geometric design. The data source used in this study was the Highway Safety Information System (HSIS), a highway safety data base administered by the Federal Highway Administration (FHWA) (10).

POISSON REGRESSION MODEL

The Poisson regression model employed in this paper was proposed by Miaou et al. (8) to develop the relationship between vehicle accidents and highway geometric design. In theory, the model can be applied to any roadway class, vehicle configuration, and accident severity type of interest. The following presentation focuses on accidents of all severity types involving large trucks (more than 4,545 kg or 10,000 lb) on a particular roadway class.

The Model

Consider a set of n road sections of a particular roadway type; for example, a rural Interstate. Let Y_i be a random variable representing the number of trucks involved in accidents on road section i during a period of 1 year, where $i = 1, 2, \ldots, n$. Here the same road section in different sample periods can be considered as separate road sections. This allows the year-to-year changes in geometric design and traffic conditions to be considered in the model. Further, the actual observation of Y_i during the period is denoted as y_i, where $y_i = 0, 1, 2, 3, \ldots$ and $i = 1, 2, \ldots, n$. The amount of truck travel (or truck exposure) during the sample year on road section i, denoted by v_i, is computed as

$$365 \times \text{AADT}_i \times (T\%/100) \times l_i$$

where

AADT_i = average annual daily traffic (in number of vehicles),

$T\%_i$ = average percentage of trucks (or percent trucks) in the traffic stream (e.g., 15), and

l_i = length (in km or mi) of road section i.

Note that $\text{AADT}_i \times (T\%_i/100)$ represents the "truck AADT" of road section i during the year. Associated with each road section i, there is a $k \times 1$ covariate vector, x_i, describing its geometric design characteristics, traffic conditions, and other relevant attributes. The transpose of the covariate vector is denoted by $x_i' = (x_{i1}, x_{i2}, \ldots, x_{ik})$. Without loss of generality, let the first covariate x_{i1} be a dummy variable equal to one for all i (i.e., $x_{i1} = 1$).

To the extent possible, these n road sections should be selected to cover as much variation in geometric design, traffic conditions, and other relevant attributes as possible. In addition, to avoid the bias in estimating the truck accident-geometric design relationship, the selection of road sections should not be based on the outcomes of the dependent variable (i.e., y_is).

Under the assumption that (a) truck exposure data and other covariates are free from errors, (b) the occurrences of truck accidents on different road sections are independent, and (c) the number of trucks involved in accidents on a particular road section i, Y_i, follows a Poisson distribution, Miaou et al. (8) proposed the following model to establish the relationship between truck accidents and highway geometric design:

$$p(Y_i = y_i) = p(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!},$$

$$y_i = 0, 1, 2, 3, \ldots, i = 1, 2, 3, \ldots, n, \tag{1}$$

where

$$\mu_i = \text{E}(Y_i) = v_i[e^{\underline{x}_i'\underline{\beta}}]$$

$$= v_i[e^{\Sigma_{j=1}^{k} x_{ij}\beta_j}] = v_i\lambda_i \quad i = 1, 2, 3, \ldots, n. \tag{2}$$

and $\underline{\beta}$ is a $k \times 1$ vector of unknown regression coefficients to be estimated from the data, the transpose of which is denoted by $\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$. This model assumes that the number of trucks involved in accidents Y, $i = 1, 2, \ldots, n$, are independently and Poisson distributed with mean μ_i, and the mean μ_i (i.e., the expected number of trucks involved in accidents) is proportional to truck travel v_i. The model also assumes an exponential rate function, $\lambda_i = \text{E}(Y_i)/v_i = \exp(\underline{x}_i'\underline{\beta})$, which ensures that accident involvement rate is always nonnegative. This type of rate function has been widely employed in statistical literature and found to be very flexible in fitting different types of count data (11,12). Note that whenever appropriate, higher order and interaction terms of covariates can be included in Equation 2 without difficulty.

On the basis of the model, the variance, $Var(Y_i)$, and coefficient of skewness, $skew(Y_i)$, of the underlying distribution of Y_i are μ_i and $\mu_i^{-1/2}$, respectively. The variance $Var(Y_i)$, which is equal to the mean μ_i, depends on its rate function and thus involves unknown regression coefficients. In addition, $Var(Y_i)$ grows linearly with truck exposure v_i. The model supposes a positive skewness coefficient that varies from road section to road section, depending on their means $[skew(Y_i) = \mu_i^{-1/2}]$. As mean μ_i increases, either as a result of an increase in vehicle exposure v_i or an increase in the rate function λ_i, the skewness coefficient $skew(Y_i)$ decreases, and as μ_i decreases, $skew(Y_i)$ increases.

Estimation and Statistical Inference

In this paper, the regression coefficients of the Poisson regression model are estimated using the maximum likelihood method. The maximum likelihood estimates (MLE) of the regression coefficients are obtained by maximizing the following loglikelihood function:

$$
\begin{aligned}
L(\underline{\beta}) &= \log\left(\prod_{i=1}^{n} p(y_i)\right) = \log\left(\prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}\right) \\
&= \sum_{i=1}^{n} [y_i \log(\mu_i) - \mu_i - \log(y_i!)] \\
&= \sum_{i=1}^{n} [y_i(\underline{x}_i'\underline{\beta}) + y_i \log(v_i) - v_i e^{\underline{x}_i'\underline{\beta}} - \log(y_i!)] \qquad (3)
\end{aligned}
$$

The first derivative of the loglikelihood function with respect to the jth regression coefficient can be shown to be

$$\frac{\partial L(\underline{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \{y_i - v_i e^{\underline{x}_i'\underline{\beta}}\} x_{ij} \qquad (4)$$

where $j = 1, 2, \ldots, k$ and must all vanish at the MLE $\hat{\underline{\beta}}$. Because the first covariate x_{i1} is a dummy variable equal to 1 for all i, the MLE requires that $\Sigma_i y_i = \Sigma_i v_i \exp(\underline{x}_i'\hat{\underline{\beta}})$. That is, the (estimated) expected total number of accident involvements, $\Sigma_i \hat{\mu}_i$, has to be equal to the observed total $\Sigma_i y_i$, where $\hat{\mu}_i = v_i \exp(\underline{x}_i'\hat{\underline{\beta}})$. This is a desirable statistical property in modeling vehicle accidents (*6*). Note that most of the suggested conventional multiple linear regression models for establishing geometric design-vehicle accident relationships do not have such a property (*6*).

The asymptotic covariance and t-statistics of the estimated coefficients, as the sample size n becomes infinite, can be determined using the second derivative of the loglikelihood function (i.e., Fisher's information matrix), as follows. The second derivative, or the Hessian matrix, of the loglikelihood function can be derived as

$$
\begin{aligned}
h_{jq} &= \frac{\partial^2 L(\underline{\beta})}{\partial \beta_j \partial \beta_q} = -\sum_i (v_i e^{\underline{x}_i'\underline{\beta}}) x_{ij} x_{iq} \\
j &= 1, 2, \ldots, k,\ q = 1, 2, \ldots, k \qquad (5)
\end{aligned}
$$

which is a function of unknown regression coefficient $\underline{\beta}$, and does not involve dependent variable y_i. Provided the Poisson assumption is adequate and the sample size is reasonably large, the asymptotic covariance matrix of the MLE can be obtained as

$$cov(\hat{\underline{\beta}}) = [I(\hat{\underline{\beta}})]^{-1} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ s_{k1} & s_{k2} & \cdots & s_{kk} \end{bmatrix} \qquad (6)$$

where

$$I(\hat{\underline{\beta}}) = -\left.\frac{\partial^2 L(\underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}'}\right|_{\underline{\beta}=\hat{\underline{\beta}}} = -\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ h_{k1} & h_{k2} & \cdots & h_{kk} \end{bmatrix}_{\underline{\beta}=\hat{\underline{\beta}}} \qquad (7)$$

is the Fisher information matrix evaluated at the MLE $\hat{\underline{\beta}}$ (*13*). The asymptotic t-statistic for each estimated regression coefficient $\hat{\beta}_j$ is computed as $\hat{\beta}_j/(s_{jj})^{1/2}$, where $(s_{jj})^{1/2}$ is the asymptotic standard deviation of $\hat{\beta}_j$ and $j = 1, 2, \ldots, k$, and its significance level can be assessed using a t distribution table with n-k degrees of freedom or simply using a normal probability table because of large n. The asymptotic correlation matrix of the estimated regression coefficients can be constructed as $\hat{\rho}_{ij} = s_{ij}/(s_{ii}s_{jj})^{1/2}$, for $i = 1, 2, \ldots, k$, and $j = 1, 2, \ldots, k$. (Note that $\hat{\rho}_{ii} = 1$ for $i = 1, 2, \ldots, k$.)

A limitation of using the Poisson regression model, which is well known in the statistical literature (*14,15*), is that the variance of the data is restrained to be equal to the mean. In many applications, count data were found to display extra variation or overdispersion relative to a Poisson model (*15*). That is, the variance of the data was greater than the Poisson model indicated.

If the overdispersion exists in the data, the MLE of the regression coefficients, $\hat{\underline{\beta}}$, under the Poisson regression model, will still be close to the true coefficients, $\underline{\beta}$, when the sample size n is large. (This is assuming that the rate function in Equation 2 has the correct form.) However, under the Poisson regression model, the variances of the estimated coefficients (i.e., s_{jj}, $j = 1, 2, \ldots, k$) would tend to be underestimated and, therefore, the associated t-statistics $\hat{\beta}_j/(s_{jj})^{1/2}$, $j = 1, 2, \ldots, k$, would tend to be overestimated (*16*). Following Wedderburn (*17*), to correct for the overdispersion problem for the Poisson regression model, it can be assumed that the variance of Y_i is $\tau\mu_i$ instead of μ_i, as originally assumed in the Poisson model, where τ is called the overdispersion parameter (typically, $\tau \geq 1$). Furthermore, a moment estimator of the overdispersion parameter τ is $\hat{\tau} = X^2/(n - k)$, where X^2 is the Pearson's chi-square statistic, n is the number of observations (i.e., the number of road sections in this case), and k is the number of unknown regression coefficients in the Poisson regression model. The Pearson's X^2 statistic is computed as $\Sigma_i(y_i - \hat{\mu}_i)^2/\hat{\mu}_i$. A better estimate of the asymptotic covariance matrix of the estimated coefficients is $\hat{\tau} \times cov(\hat{\underline{\beta}})$ and, therefore, a better estimate of the t-statistic for regression coefficient $\hat{\beta}_j$ is $\hat{\beta}_j/(\hat{\tau}\, s_{jj})^{1/2}$, $j = 1, 2, \ldots, k$ [see, e.g., Agresti (*18*)].

Model Applications

To illustrate how the Poisson regression model can be used to estimate the expected reduction in truck accident involvements caused by improvements in some geometric design elements, consider a particular road section i, and let the value of its covariates before and after the improvement be x_{ij}^b and x_{ij}^a, for $j = 1, 2, \ldots, k$. Also, let v_i^b and v_i^a be the amount of truck travel in one year on road section i before and after the improvement.

Based on the Poisson regression model (Equations 1 and 2), the expected number of truck accident involvements on road section i before and after the improvements of geometric design elements are, respectively, $v_i^b \exp(\Sigma_j x_{ij}^b \beta_j)$ and $v_i^a \exp(\Sigma_j x_{ij}^a \beta_j)$. The percentage reduction in the expected truck accident involvements can be computed as

$$R_i = \left[\frac{v_i^b \exp\left(\sum_{j=1}^{k} x_{ij}^b \beta_j\right) - v_i^a \exp\left(\sum_{j=1}^{k} x_{ij}^a \beta_j\right)}{v_i^b \exp\left(\sum_{j=1}^{k} x_{ij}^b \beta_j\right)} \right] \times 100$$

$$= \left\{ 1 - \left(\frac{v_i^a}{v_i^b}\right) \exp\left[\sum_{j=1}^{k} (x_{ij}^a - x_{ij}^b)\beta_j\right] \right\} \times 100 \qquad (8)$$

The percentage reduction R_i is sometimes referred to as the truck accident involvement reduction factor. If v_i is the same before and after the improvement (i.e., $v_i^b = v_i^a$) then R_i also represents the percentage reduction in truck accident involvement rate. By substituting β_j with the MLE $\hat{\beta}_j$ in Equation 8 for $j = 1, 2, \ldots, k$, a MLE of the reduction in the expected number of truck accident involvements can be obtained, denoted by $\hat{R}_i$. Because, for a large sample, $\underline{\hat{\beta}}$ is approximately normally distributed with mean $\underline{\beta}$ and with covariance matrix $\hat{\tau} \times cov(\underline{\hat{\beta}})$ [see, for example, Agresti (*18*)], it can be shown that the standard deviation ($s.d.$) of $\hat{R}_i$ is approximately as follows:

$$s.d.(\hat{R}_i) \approx \left(\frac{v_i^a}{v_i^b}\right) \times \left\{ \exp\left[\sum_{j=1}^{k} (x_{ij}^a - x_{ij}^b)\hat{\beta}_j + \frac{\hat{\tau}}{2} \sum_{m=1}^{k} \sum_{q=1}^{k} \right.\right.$$
$$\left.\left. \times (x_{im}^a - x_{im}^b)(x_{iq}^a - x_{iq}^b)\hat{\rho}_{mq}(s_{mm}s_{qq})^{1/2} \right]\right\}$$
$$\times \left\{ \exp\left[\hat{\tau} \sum_{m=1}^{k} \sum_{q=1}^{k} (x_{im}^a - x_{im}^b) \right.\right.$$
$$\left.\left. \times (x_{iq}^a - x_{iq}^b)\hat{\rho}_{mq}(s_{mm}s_{qq})^{1/2} \right] - 1 \right\}^{1/2} \times 100 \qquad (9)$$

The derivation uses the property that if z is normally distributed with mean μ and variance σ^2, then the variance of $\exp(z)$ is $\{\exp[\mu + (1/2)\sigma^2]\}^2 [\exp(\sigma^2) - 1]$ [see Lindgren, (*19*), page 191]. Equation 9 allows the uncertainty of the estimated reduction to be assessed by quoting plus or minus one standard deviation.

ILLUSTRATION AND DISCUSSION

Data Source

To illustrate the use of the Poisson regression model, data from the HSIS were employed to develop relationships between truck accidents and key highway geometric design variables. The HSIS currently has data from five states. A general description of the HSIS data base is given in Council and Paniati (*10*). Specifically, accidents involving trucks (of more than 4,545 kg or 10,000 lb) on rural Interstate highways from Utah were used. Among the five HSIS States, Utah was considered to be the state that had the most complete information on highway geometric design. In addition, Utah was the only HSIS state with a historical road inventory file in which year-to-year changes on highway geometric design element and traffic conditions were recorded. Thus, accidents in a given year could be matched to the road inventory information of the same period. Data from 1985 to 1989 were used for the illustration.

Utah data in HSIS were stored in six files: roadlog, horizontal curvature, vertical grade, accident, vehicle, and occupant files. Thus, these files had to be linked before any analysis could be performed. Each record in the roadlog file represented a homogeneous section in terms of its cross-sectional characteristics, such as number of lanes, lane width, shoulder width, median type and width, AADT, and percent trucks. However, these road sections were not necessarily homogeneous in terms of their horizontal curvatures and vertical grades. Road sections in the horizontal curvature and vertical grade files, on the other hand, were delineated in such a way that they were homogeneous in terms of their horizontal curvatures and vertical grades, respectively, but not necessarily in terms of other road characteristics.

Therefore, after matching road sections in the horizontal curvature and vertical grade files with the road sections in the roadlog file, each road section in the road inventory file may have contained more than one horizontal curvature or vertical grade. In this illustration, those road sections with multiple curvatures and grades were further disaggregated into smaller subsections so that each subsection contained a unique set of horizontal curvature and vertical grade. Each subsection, which was totally homogeneous in cross-sectional characteristics, horizontal curvature, and vertical grade, was then considered as an independent road section in the model. In order to test the effects of the length of curve and grade, information on the length of the original curve and grade, from which the subsection was delineated, was maintained for each subsection.

Accidents, Characteristics of Road Sections, and Covariates

The time period considered was 1 year, which means that the same road section, even if nothing had changed, was considered as five independent sections—one for each year from 1985 to 1989. As indicated earlier, this allowed the year-to-year changes on highway geometric design and traffic conditions to be considered in the model. A total of 8,263 homogeneous road sections during the 5-year period were considered to have reliable data. These road sections covered about 99 percent of the entire rural Interstate highway mileage in Utah and constituted 23,570 lane-km or 14,731 lane-mi of roadway. Data for each year contained roughly ⅕ of the total sections and lane-km. The section lengths varied from 0.016 to 12.43 km (0.01 to 7.77 mi)—with an average of 0.72 km (0.45 mi). Descriptive statistics of these 8,263 road sections on truck accident involvements and truck miles (km) traveled are given in Table 1.

TABLE 1 Variable Definitions and Summary Statistics of the 8,263 Rural Interstate Road Sections

Variable	Notation & Definition (for section i)	Min	Max	Mean	% Zero
Number of Trucks Involved in Accidents	y_i	0	8	0.20	86
Section Length (in mi)	ℓ_i	0.01	7.77	0.45	0
Truck Miles or Truck Exposure (in 10^6 truck-miles)	v_i = [365×AADT$_i$×(T%$_i$/100)×ℓ_i]/10^6, where T%$_i$ is percent trucks (366 for leap years).	8×10^{-4}	5.03	0.25	0
Dummy Intercept	$x_{i1} = 1$				
Dummy Variable for Year 1986, representing year-to-year changes due to random fluctuations, annual trend, and omitted variables such as weather.	x_{i2} = 1, if the road section is in year 1986 = 0, otherwise				
Dummy Variable for Year 1987 (See above explanation)	x_{i3} = 1, if the section is in 1987 = 0, otherwise				
Dummy Variable for Year 1988 (See above explanation)	x_{i4} = 1, if the section is in 1988 = 0, otherwise				
Dummy Variable for Year 1989 (See above explanation)	x_{i5} = 1, if the section is in 1989 = 0, otherwise				
AADT per Lane (in 1000's of vehicles), a surrogate variable to indicate traffic conditions or traffic density.	x_{i6} = (AADT$_i$/number of lanes$_i$)/1000	0.35	12.04	1.80	0
Horizontal Curvature, HC, (in degrees per 100-ft arc)	x_{i7}	0	12.00	1.00	67
Length of Original Horizontal Curve, LHC, (in mi) from which this curve was subdivided for creating homogeneous sections; only for HC > 1 and LHC ≤ 1.	x_{i8} = LHC, if $x_{i7}>1$ and LHC ≤ 1 mi. = 1.0, if $x_{i7}>1$ and LHC > 1 mi. = 0, if $x_{i7}\le1$	0	0.96	0.05	81
Vertical Grade, VG, (in percent)	x_{i9}	0	8.00	2.14	20
Length of Original Vertical Grade, LVG, (in mi) from which this section was subdivided for creating homogeneous sections; only for sections with VG > 2 and LVG ≤ 2.	$x_{i,10}$ = LVG, if $x_{i9}>2$ and LVG ≤ 2 mi. = 2.0, if $x_{i9}>2$ and LVG > 2 mi. = 0, if $x_{i9}\le2$	0	2.00	0.21	74
Deviation of Paved Inside Shoulder Width (per direction) from an "ideal" width of 12 ft (3.66 m).	$x_{i,11}$ = max{0, 12 - paved inside shoulder width}	4.00	12.00	8.16	0
Percent Trucks in the traffic stream (e.g., 15)	$x_{i,12}$	7.00	57.00	24.13	0
HC × LHC	$x_{i,13} = x_{i7} \times x_{i8}$	0	2.88	0.18	81
VG × LVG	$x_{i,14} = x_{i9} \times x_{i,10}$	0	13.37	0.97	74

(1 mi = 1.61 km; 1 ft = 0.3048 m)

During the 5-year period, 1,643 large trucks were reported to be involved in accidents on these highway sections, regardless of truck configuration and accident severity type. With the total truck travel estimated to be 3,248 million truck km (MTK) or 2,030 million truck mi (MTM), the overall truck accident involvement rate was therefore 0.51 truck accident involvements/MTK or 0.81 truck accident involvements/MTM. These accidents occurred on only 14 percent of the 8,263 road sections. The maximum number of trucks involved in accidents on an individual road section in one year was 8. On average, each section had 0.20 trucks involved in accidents in 1 year.

The covariates considered for individual road sections and their definitions are also presented in Table 1. They include

1. Yearly dummy variables to capture year-to-year changes in the overall truck accident involvement rate caused by, for example, long-term trend, annual random fluctuations, changes in posted speed limit, and changes in omitted variables such as weather;
2. AADT/lane, used as a surrogate measure for traffic flow density;
3. Horizontal curvature (HC);
4. Vertical grade (VG); and
5. Deviation of paved inside (or left) shoulder width/direction from an "ideal" width of 3.66 m (12 ft).

Note that paved inside shoulder width (ISH) of 3.66 m (12 ft)/direction is recommended by the "Greenbook" for roads with heavy truck traffic (*20*). Because all of the road sections were 3.66 m (12 ft) in lane width, more than 89 percent of them had 4 lanes, and all road sections had paved outside (or right) shoulder widths of 3.05 m (10 ft), the effects of these variables could not be determined in this study.

It has been suggested that as the length of grade increases to a point that can slow a truck to a speed significantly slower than the speed of the traffic stream (e.g., 16 km/hr or 10 mi/hr), the accident rate increases (*3*). Also, for a fixed curvature degree, as the length of curve increases, the accident rate increases (*21*). To test the effects of length of curve and length of grade on truck accident involvement rate, two covariates—length of original curve (LHC), $x_{i,8}$, and length of original grade (LVG), $x_{i,10}$,—were considered. As indicated earlier, each curve or grade considered in the model may have been subdivided from a longer curve or grade for achieving total homogeneity. Thus, for each road section in the model, these two covariates were defined as the length of the original undivided curve or undivided grade to which this section belonged. In addition, these two covariates were defined only for curves with horizontal curvatures greater than 1 degree per 30.48-m (100-ft) arc and sections with grade greater than 2 percent. (Note that these two covariates were set equal to 0 if horizontal curvature is less than or equal to 1 degree or if vertical grade is less than or equal to 2 percent.) This definition was based on an assumption that the length of a mild curve or grade has no aggravated effect on truck accident involvement rate. On the basis of engineering judgments, it was further assumed that there were no additional effects after LHC reached 1.6 km (1.0 mi) or after LVG reached 3.2 km (2.0 mi). This assumption makes the effects of LHC and LVG on truck accident involvement rate more robust to unusually long curves and grades. The interactions of HC and LHC ($x_{i,13} = x_{i,7} \times x_{i,8}$), VG and LVG ($x_{i,14} = x_{i,9} \times x_{i,10}$), and HC and VG ($x_{i,7} \times x_{i,9}$) were also considered.

Percent trucks in the traffic stream was included in the model to evaluate the effects of automobile-truck mix. Previous studies suggested that as percent trucks increases, truck accident involvement rate decreases. One possible reason is that, for a constant vehicle density, as percent trucks increases, the frequency of lane changing and overtaking movements by automobiles decreases. Also, previous records showed that more trucks were involved in truck-automobile multivehicle accidents than in truck-truck accidents [e.g., see Jovanis and Chang (*5*)].

Model Results

The estimated regression coefficients of some of the tested models using the 8,263 homogeneous road sections and the associated *t*-statistics are presented as Models 0–7 in Table 2. The estimated overdispersion parameter ($\hat{\tau}$), loglikelihood function evaluated at the estimated coefficients, $L(\hat{\beta})$, and the Akaike Information Criterion (AIC) value (*22*) for each model are also given in the table. Note that AIC $= -2L(\hat{\beta}) + 2k$, where k is the total number of regression coefficients in the model, and the estimated models with high loglikelihood function and low AIC values are preferred. Furthermore, the expected total number of trucks involved in accidents across road sections ($\Sigma_i \hat{\mu}_i$) was compared with the observed total ($\Sigma_i y_i$).

These eight models in Table 2 are arranged as follows.

Model 0: This is the simplest form of the Poisson regression model, which includes only truck exposure (v_i). That is, Y_i is assumed to be Poisson distributed with mean $\mu_i = v_i \exp(\beta_1)$. This model served as a baseline for the measurement of model improvement as additional explanatory variables were included.

Model 1: This model includes only truck exposure and yearly dummy variables (x_{ij}, $j = 2, \ldots, 5$) to capture year-to-year changes in the overall truck accident involvement rate.

Model 2: This model includes truck exposure, yearly dummy variables, and traffic variables, including AADT per lane (x_{i6}) and percent trucks ($x_{i,12}$).

Models 3–5: These models include truck exposure, yearly dummy variables, traffic variables, and geometric design variables, including horizontal curvature, length of original curve, vertical grade, length of original grade, and paved inside shoulder width (x_{ij}, $j = 7, 8, 9, 10, 11$). The interactions between horizontal curvature and length of curve ($x_{i,13}$) and between vertical grade and length of grade ($x_{i,14}$) were also tested. (Note that the interaction between HC and VG was not found to be significant at a 20 percent α level.)

Model 6: This model uses the same explanatory variables as in Model 5. It was intended for examining the effect of short road sections on the estimation of model coefficients. Only road sections with length greater than 0.08 km (0.05 mi) were used to estimate model coefficients. There were 7,004 road sections and 1,603 reported truck accident involvements.

Model 7: This model has the same explanatory variables as in Model 5. It was used for checking the effect of road sections with large model residuals on the estimation of model coefficients. Based on Model 5, there were 53 road sections with large standardized residual values [defined as road sections with $|y_i - \hat{\mu}_i|/(\hat{\tau}\ \hat{\mu}_i)^{1/2} > 5$]. These road sections were first removed and Model 5 was then recalibrated to obtain Model 7.

The following observations can be made from these eight models:

1. The AIC value continues to decrease and $L(\hat{\beta})$ continues to increase from Model 0 to Model 3, as yearly dummy variables, traffic variables, and geometric design variables are included in the model.
2. By comparing Model 3 with Model 4, it can be observed that it is through the interaction with horizontal curvature that length of curve becomes a significant factor in affecting truck accident involvement rate. This is shown by the unadjusted *t*-statistic of 0.02 for $\hat{\beta}_8$ in Model 3 and of 2.76 for $\hat{\beta}_{13}$ in Model 4.
3. It is suggested from Model 4 that length of grade by itself is a significant determinant for truck accident involvement rate. [The adjusted *t*-statistics for $\hat{\beta}_{10}$ is $1.99/(1.57)^{1/2} = 1.59$.] As can be seen from Model 5 through the interaction with vertical grade, the effect of length of grade becomes more significant. [The adjusted *t*-statistics for $\hat{\beta}_{14}$ is $2.26/(1.57)^{1/2} = 1.80$.] In this study, Model 5, which had the lowest AIC value, was considered for further analyses and illustrations. The asymptotic correlation matrix, $\hat{\rho}_{ij}$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, k$, for the estimated regression coefficients in Model 5 is shown in Table 3.
4. The comparison of the estimated coefficients of Model 5 and Model 6 suggested not only that the conclusions reached regarding the significance level of the relationships between truck accidents and the examined traffic and highway geometric variables were consistent, but also that the estimated coefficient values were very close. This suggests that the Poisson regression model is not sensitive to the length of road sections.
5. The comparison of the estimated regression coefficients for the traffic and geometric design variables (i.e., β_6 through β_{14}) between Model 5 and Model 7 suggested that the deletion

TABLE 2 Estimated Regression Coefficients of Some Tested Poisson Regression Models and Associated Statistics

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Section length and number of road sections	≥0.01 mi 8,263	≥0.01 mi 8,263	≥0.01 mi 8,263	≥0.01 mi 8,263	≥0.01 mi 8,263	≥0.01 mi 8,263	>0.05 mi 7,004	≥0.01 mi 8,210
β_1 Dummy intercept	-0.212230 (±0.025;-8.60)	0.121263 (±0.058;2.08)	0.570141 (±0.112;5.07)	-0.472330 (±0.287;-1.65)	-0.472494 (±0.287;-1.65)	-0.431762 (±0.288;-1.50)	-0.526103 (±0.290;-1.81)	-0.500055 (±0.298;-1.68)
β_2 Dummy variable for 1986	-----	-0.363320 (±0.082;-4.44)	-0.163271 (±0.086;-1.90)	-0.182576 (±0.086;-2.12)	-0.185384 (±0.086;-2.15)	-0.183853 (±0.086;-2.14)	-0.171759 (±0.087;-1.97)	-0.252277 (±0.089;-2.83)
β_3 Dummy variable for 1987	-----	-0.340802 (±0.080;-4.24)	-0.139415 (±0.085;-1.64)	-0.160249 (±0.085;-1.89)	-0.162656 (±0.085;-1.91)	-0.161461 (±0.085;-1.90)	-0.160869 (±0.086;-1.86)	-0.186897 (±0.087;-2.14)
β_4 Dummy variable for 1988	-----	-0.327909 (±0.078;-4.21)	-0.090187 (±0.085;-1.06)	-0.114524 (±0.085;-1.35)	-0.112753 (±0.085;-1.33)	-0.111511 (±0.085;-1.31)	-0.096243 (±0.086;-1.12)	-0.166516 (±0.088;-1.90)
β_5 Dummy variable for 1989	-----	-0.518223 (±0.079;-6.54)	-0.289009 (±0.088;-3.29)	-0.315484 (±0.088;-3.57)	-0.313863 (±0.088;-3.57)	-0.311155 (±0.088;-3.54)	-0.299701 (±0.089;-3.36)	-0.355124 (±0.091;-3.92)
β_6 AADT per lane (10^3)	-----	-----	0.027600 (±0.015;1.85)	0.026710 (±0.015;1.73)	0.022138 (±0.015;1.38)	0.024400 (±0.015;1.59)	0.025220 (±0.015;1.63)	0.030559 (±0.016;1.94)
β_7 Horizontal curvature	-----	-----	-----	0.147259 (±0.022;6.85)	0.089178 (±0.028;3.15)	0.088861 (±0.028;3.14)	0.096170 (±0.029;3.27)	0.081928 (±0.030;2.75)
β_8 Length of original curve	-----	-----	-----	0.004148 (±0.232;0.02)	-----	-----	-----	-----
β_{13} (Horizontal curvature)× (Length of original curve)	-----	-----	-----	-----	0.232377 (±0.084;2.76)	0.234209 (±0.084;2.78)	0.221877 (±0.087;2.56)	0.239432 (±0.088;2.73)
β_9 Vertical grade	-----	-----	-----	0.083423 (±0.027;3.06)	0.084194 (±0.027;3.09)	0.077815 (±0.028;2.81)	0.078218 (±0.028;2.78)	0.050211 (±0.028;1.77)
β_{10} Length of original grade	-----	-----	-----	0.165342 (±0.078;2.11)	0.156212 (±0.078;1.99)	-----	-----	-----
β_{14} (Vertical grade)× (Length of original grade)	-----	-----	-----	-----	-----	0.033973 (±0.01;2.26)	0.031085 (±0.015;2.03)	0.044749 (±0.015;2.89)
β_{11} Deviation of paved inside shoulder width from 12 ft	-----	-----	-----	0.088652 (±0.036;2.46)	0.091478 (±0.036;2.54)	0.085763 (±0.036;2.37)	0.094814 (±0.036;2.60)	0.088546 (±0.037;2.36)
β_{12} Percent trucks (e.g., 15)	-----	-----	-0.028940 (±0.004;-6.96)	-0.025260 (±0.004;-5.91)	-0.025738 (±0.004;-6.01)	-0.025233 (±0.004;-5.88)	-0.025308 (±0.004;-5.82)	-0.022769 (±0.004;-5.11)
τ	1.90	1.84	1.76	1.57	1.57	1.57	1.32	0.97
$L(\hat{\beta})$	-3916.4	-3895.0	-3845.5	-3775.3	-3771.7	-3771.0		
AIC Value	7834.7	7800.0	7705.0	7574.5	7567.3	7566.0		
Expected vs. Observed Total Truck Accident Involvements	1,641.8 1,643.0	1,645.6 1,643.0	1,641.6 1,643.0	1,642.3 1,643.0	1,644.2 1,643.0	1,644.3 1,643.0	1,604.5 1,603.0	1,540.8 1,539.0

Notes: (1) Values in parentheses are (unadjusted) asymptotic standard deviation and t-statistics of the coefficients above.
(2) ----- Not included in the model.
(3) 1 mile = 1.61 km, 1 ft = 0.3048 m.

TABLE 3 Asymptotic Correlation Matrix, ($\hat{\rho}_{ij}$), of the Estimated Regression Coefficients, $\hat{\beta}$, for Model 5

	1	2	3	4	5	6	7	13	9	14	11	12
1	1.000	-0.095	-0.087	-0.083	-0.054	-0.193	-0.059	0.034	-0.080	0.099	-0.905	-0.058
2	-0.095	1.000	0.564	0.586	0.578	-0.169	-0.026	0.017	-0.001	-0.029	0.059	-0.305
3	-0.087	0.564	1.000	0.601	0.594	-0.194	-0.029	0.018	0.004	-0.034	0.058	-0.324
4	-0.083	0.586	0.601	1.000	0.628	-0.256	-0.037	0.025	0.012	-0.048	0.080	-0.401
5	-0.054	0.578	0.594	0.628	1.000	-0.314	-0.039	0.029	0.013	-0.053	0.070	-0.425
6	-0.193	-0.169	-0.194	-0.256	-0.314	1.000	0.047	-0.076	-0.070	0.127	-0.076	0.590
7	-0.059	-0.026	-0.029	-0.037	-0.039	0.047	1.000	-0.792	-0.050	0.024	0.009	0.106
13	0.034	0.017	0.018	0.025	0.029	-0.076	-0.792	1.000	-0.007	-0.020	-0.003	-0.069
9	-0.080	-0.001	0.004	0.012	0.013	-0.070	-0.050	-0.007	1.000	-0.783	-0.068	0.001
14	0.099	-0.029	-0.034	-0.048	-0.053	0.127	0.024	-0.020	-0.783	1.000	-0.041	0.113
11	-0.905	0.059	0.058	0.080	0.070	-0.076	0.009	-0.003	-0.068	-0.041	1.000	-0.281
12	-0.058	-0.305	-0.324	-0.401	-0.425	0.590	0.106	-0.069	0.001	0.113	-0.281	1.000

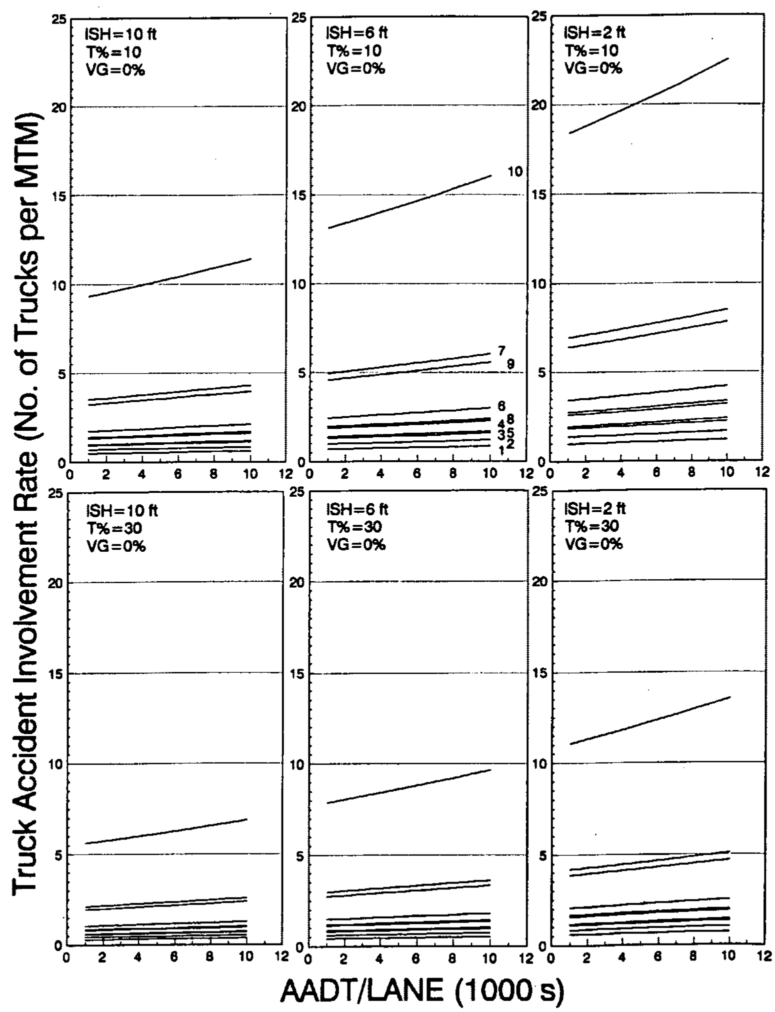


FIGURE 1 The relationship between truck accident involvement rate and key highway geometric design variables for rural Interstate highways *(continued on next page)*.

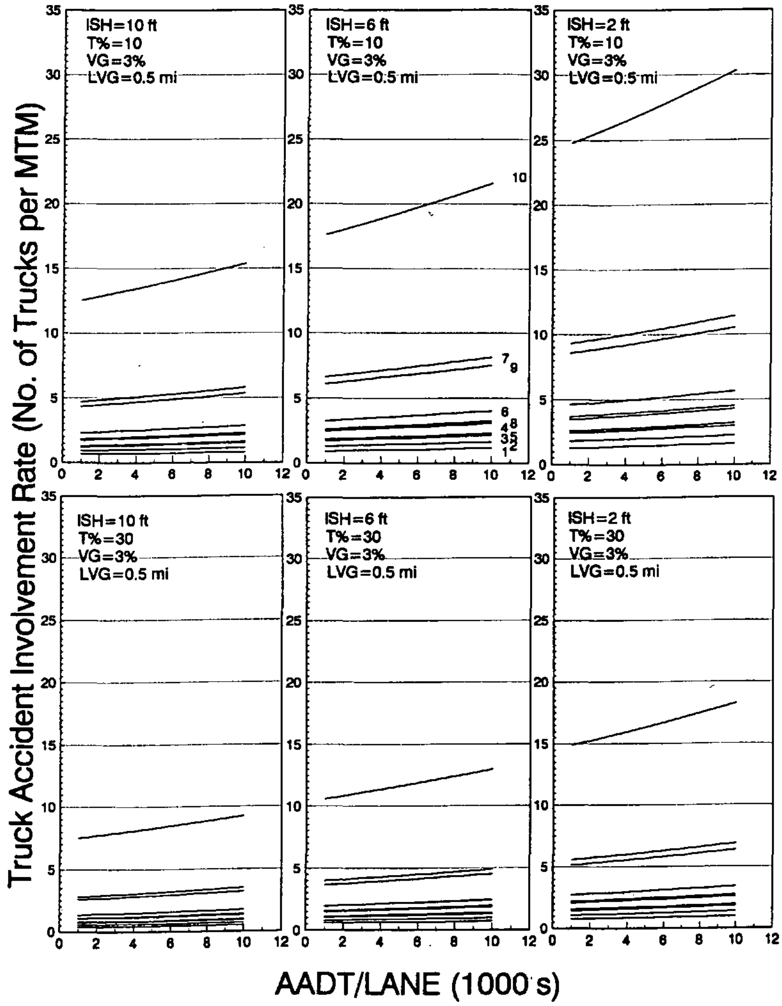


FIGURE 1 (*continued*).

of the 53 road sections with high standardized residuals only slightly altered the coefficient estimates. This also meant that no particular road section had unusually high influence on the estimates.

6. All of the estimated coefficients for the traffic and geometric variables are consistent among different models and have expected algebraic signs.

7. Based on Model 5, truck accident involvement rates for different combinations of AADT/lane, horizontal curvature, length of original curve, vertical grade, length of original grade, paved inside shoulder width, and percent trucks are illustrated in Figure 1. These rates are computed using the average estimated coefficients for 1987–1989 dummy variables as a base rate: $\hat{\lambda}_i = \exp[\hat{\beta}_1 + (\hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5)/3 + x_{i,6}\hat{\beta}_6 + x_{i,7}\hat{\beta}_7 + x_{i,13}\hat{\beta}_{13} + x_{i,9}\hat{\beta}_9 + x_{i,14}\hat{\beta}_{14} + x_{i,11}\hat{\beta}_{11} + x_{i,12}\hat{\beta}_{12}] = \exp[-0.626471 + 0.02440x_{i,6} + 0.088861x_{i,7} + 0.234209x_{i,13} + 0.077815x_{i,9} + 0.033973x_{i,14} + 0.085763x_{i,11} - 0.025233x_{i,12}]$
[In Figure 1, Lines 1 through 10 in each part of the figure show truck accidents–geometric design relationships for different combinations of horizontal curvature (HC) in degrees per 30.48-m (100-ft arc) and length of original curve (LHC) in mi: Line 1: HC = 0; Line 2: HC = 3, LHC = 0.1; Line 3: HC = 3, LHC = 0.5; Line 4: HC = 3, LHC = 1.0; Line 5: HC = 6, LHC = 0.1; Line 6: HC = 6, LHC = 0.5; Line 7: HC = 6, LHC = 1.0; Line 8: HC = 9, LHC = 0.1; Line 9: HC = 9, LHC = 0.5; Line 10: HC = 9, LHC = 1.0. Note that this figure applies mainly to road sections with 3.66-m (12-ft) lane width and 3.05-m (10-ft) paved outside shoulder width. Also, in each part of the figure, the line numbers from the bottom to the top are: 1, 2, 3, 5, 4, 8, 6, 9, 7, and 10.]

8. For the ranges of covariates indicated in Table 1, Model 5 suggests the following relationships between geometric design elements and truck accident involvement rates:

1. As AADT/lane increases by 1,000 vehicles/lane, truck accident involvement rate increases by about 2.5 percent.

2. As horizontal curvature increases, truck accident involvement rate increases. However, the increase depends on the length of curve. For example, for a curve with 0.1 mi in length and with curvature greater than 1 degree/30.48-m (100-ft) arc, as horizontal curvature increases by 1 degree, truck accident involvement rate increases by about 11.9 percent.

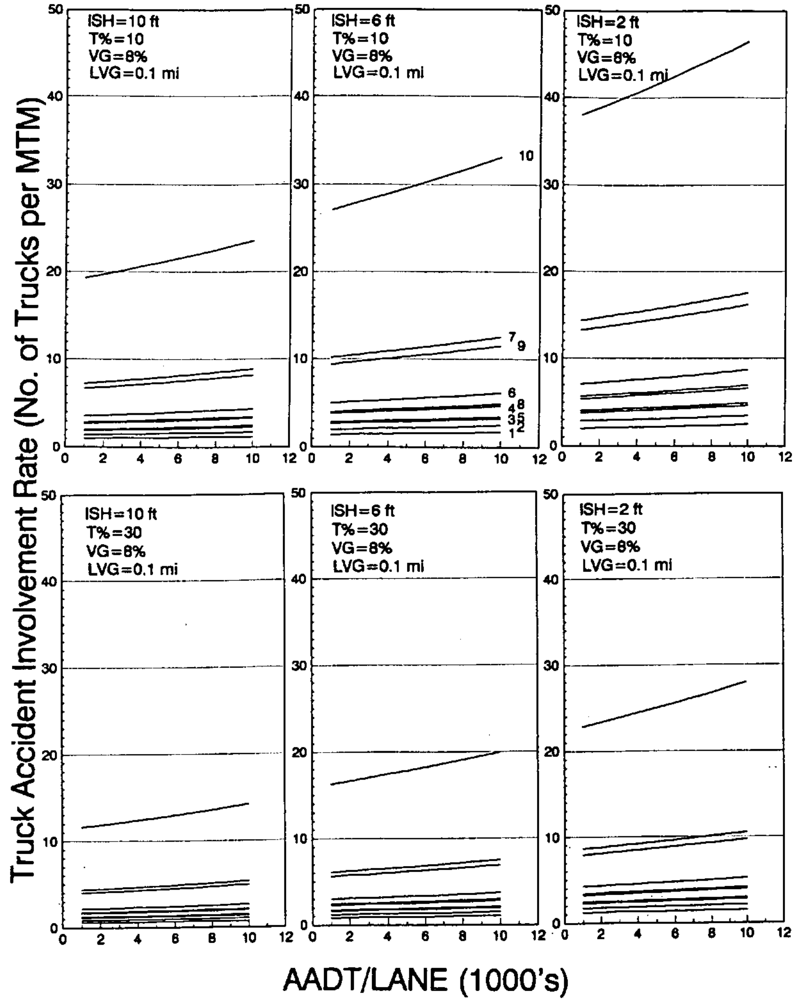


FIGURE 1 *(continued).*

3. As vertical grade increases, truck accident involvement rate increases. The increase, however, depends on the length of grade. For example, for a grade with 0.8 km (0.5 mi) in length and with vertical grade greater than 2 percent, as grade increases by 1 percent, truck accident involvement rate increases by about 9.9 percent.

4. As the length of curve increases, truck accident involvement rate increases. The increase, however, depends on the curvature degree. For example, for a 5 degree curve, as the length of curve increases by 0.16 km (0.1 mi), truck accident involvement rate increases by about 7.3 percent.

5. As the length of grade increases, truck accident involvement rate increases. The increase depends on the steepness of vertical grade. For example, for a 5 percent grade, as the length of grade increases by 0.8 km (0.5 mi), truck accident involvement rate increases by about 5.2 percent.

6. As paved inside shoulder width per direction increases by 0.3048 m (1 ft), truck accident involvement rate decreases by about 8.2 percent.

7. For a constant vehicle density, as percent trucks in the traffic stream increases by 5 percent, truck accident involvement rate decreases by about 11.9 percent.

Example Applications

Based on Model 5, the reduction in the expected number of truck accident involvements and its estimated one-standard deviation (from Equations 8 and 9) caused by improvements in horizontal curvature, vertical grade, and paved inside shoulder width of a road section, are illustrated in Tables 4 and 5. These illustrations assume no changes in truck travel after the improvements. The expected reductions caused by an improvement in one geometric design element are shown in Table 4, and the expected reductions caused by improvements in two geometric design elements are shown in Table 5. Note that Equations 8 and 9 can be used to estimate the expected reductions of a road section caused by improvements in any combination of geometric design elements.

TABLE 4 Expected Reductions in Truck Accident Involvements on a Rural Interstate Road Section After an Improvement in One Geometric Design Element

Length of Original Curve (mi)	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
0.10	10.6% (±2.5%)	20.1% (±4.5%)	28.6% (±6.0%)	36.2% (±7.2%)	43.0% (±8.1%)
0.25	13.7% (±1.9%)	25.5% (±3.3%)	35.7% (±4.2%)	44.5% (±4.9%)	52.1% (±5.3%)
0.50	18.6% (±2.7%)	33.8% (±4.4%)	46.1% (±5.4%)	56.1% (±5.8%)	64.3% (±6.0%)
0.75	23.2% (±4.3%)	41.1% (±6.6%)	54.8% (±7.7%)	65.3% (±8.0%)	73.4% (±7.8%)
≥1.00	27.6% (±5.8%)	47.6% (±8.6%)	62.1% (±9.6%)	72.5% (±9.5%)	80.1% (±9.0%)

Length of Original Grade (mi)	Vertical Grade (VG): for 2% < VG <9%				
	Reduce 1%	Reduce 2%	Reduce 3%	Reduce 4%	Reduce 5%
0.10	7.8% (±3.1%)	15.0% (±5.7%)	21.6% (±7.9%)	27.7% (±9.7%)	33.4% (±11.3%)
0.50	9.0% (±2.5%)	17.3% (±4.6%)	24.7% (±6.3%)	31.5% (±7.7%)	37.7% (±8.8%)
1.00	10.6% (±2.1%)	20.0% (±3.7%)	28.5% (±5.0%)	36.0% (±5.9%)	42.8% (±6.7%)
≥2.00	13.5% (±2.1%)	25.3% (±3.6%)	35.4% (±4.6%)	44.2% (±5.4%)	51.7% (±5.8%)

	Paved Inside Shoulder Width (ISH) per Direction: for ISH ≤ 12 ft				
	Increase 1 ft	Increase 2 ft	Increase 3 ft	Increase 4 ft	Increase 5 ft
	8.2% (±4.2%)	15.7% (±7.7%)	22.7% (±10.7%)	29.0% (±13.2%)	34.9% (±15.4%)

Notes: (1) Values in parentheses are one standard deviation of the expected reductions above.
(2) 1 ft = 0.3048 m; 1 mi = 1.61 km.

TABLE 5 Expected Reductions in Truck Accident Involvements on a Rural Interstate Road Section After an Improvement in Two Geometric Design Elements

Length of Original Curve (LHC) = 0.10 mi and Length of Original Grade (LVG) = 0.50 mi					
Vertical Grade (VG): for 2% < VG < 9%	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
Reduce 1%	18.7% (±3.1%)	27.3% (±4.4%)	35.0% (±5.6%)	42.0% (±6.6%)	48.1% (±7.4%)
Reduce 2%	26.0% (±4.5%)	33.9% (±5.0%)	40.9% (±5.8%)	47.2% (±6.4%)	52.8% (±7.0%)
Reduce 3%	32.7% (±5.8%)	39.9% (±5.9%)	46.3% (±6.2%)	52.0% (±6.5%)	57.1% (±6.9%)
Reduce 4%	38.8% (±7.0%)	45.3% (±6.7%)	51.1% (±6.7%)	56.3% (±6.7%)	61.0% (±6.9%)
Reduce 5%	44.3% (±7.9%)	50.3% (±7.4%)	55.5% (±7.1%)	60.3% (±7.0%)	64.5% (±6.9%)

Length of Original Curve (LHC) = 0.10 mi					
Paved Inside Shoulder Width per Direction (ISH): for ISH ≤ 12 ft	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
Increase 1 ft	18.0% (±4.4%)	26.7% (±5.3%)	34.5% (±6.3%)	41.4% (±7.1%)	47.6% (±7.8%)
Increase 2 ft	24.7% (±7.2%)	32.7% (±7.3%)	39.9% (±7.5%)	46.2% (±7.9%)	52.0% (±8.2%)
Increase 3 ft	30.9% (±9.8%)	38.2% (±9.3%)	44.8% (±9.0%)	50.7% (±8.9%)	55.9% (±8.9%)
Increase 4 ft	36.6% (±12.0%)	43.3% (±11.1%)	49.3% (±10.5%)	54.7% (±10.0%)	59.5% (±9.7%)
Increase 5 ft	41.8% (±13.8%)	48.0% (±12.7%)	53.5% (±11.8%)	58.4% (±11.1%)	62.8% (±10.5%)

Notes: (1) Values in parentheses are one standard deviation of the expected reductions above.
(2) 1 ft = 0.3048 m; 1 mi = 1.61 km.

To give a simple illustration of the computations involved, consider a curved road section i with 0.16 km (0.10 mi) in length. By reducing 1 degree (per 30.48-m or 100-ft arc) of the curve and all else equal, the expected truck accident involvement reduction percentage is calculated as

$$\begin{aligned}
\hat{R}_i &= \{1 - \exp[(x^a_{i,7} - x^b_{i,7})\hat{\beta}_7 + (x^a_{i,13} - x^b_{i,13})\hat{\beta}_{13}]\} \times 100 \\
&= \{1 - \exp[(x^a_{i,7} - x^b_{i,7})\hat{\beta}_7 + (x^a_{i,7} \times x^a_{i,8} - x^b_{i,7} \times x^b_{i,8})\hat{\beta}_{13}]\} \\
&\quad \times 100 \\
&= \{1 - \exp[(-1) \times 0.088861 + (-1 \times 0.1) \times 0.234209]\} \\
&\quad \times 100 \\
&= [1 - \exp(-0.1123)] \times 100 \\
&= 10.6
\end{aligned}$$

The standard error of this expected reduction percentage is computed using Equation 9 as

$$\begin{aligned}
s.d.(\hat{R}_i) \approx &\left\{\exp\left[(x^a_{i,7} - x^b_{i,7})\hat{\beta}_7 + (x^a_{i,13} - x^b_{i,13})\hat{\beta}_{13} + \frac{\hat{\tau}}{2}\right.\right. \\
&\times \left((x^a_{i,7} - x^b_{i,7})^2 s_{7,7} + (x^a_{i,13} - x^b_{i,13})^2 s_{13,13}\right. \\
&\left.\left.\left. + 2(x^a_{i,7} - x^b_{i,7})(x^a_{i,13} - x^b_{i,13})\hat{\rho}_{7,13}(s_{7,7}s_{13,13})^{1/2}\right)\right]\right\} \\
&\times \left\{\exp\left[\hat{\tau}\left((x^a_{i,7} - x^b_{i,7})^2 s_{7,7} + (x^a_{i,13} - x^b_{i,13})^2 s_{13,13}\right.\right.\right. \\
&\left.\left.\left. + 2(x^a_{i,7} - x^b_{i,7})(x^a_{i,13} - x^b_{i,13})\hat{\rho}_{7,13}(s_{7,7}s_{13,13})^{1/2}\right)\right] - 1\right\}^{1/2} \\
&\times 100 \\
= &\left\{\exp\left[-0.1123 + \frac{1.57}{2}\left((-1)^2(0.028)^2\right.\right.\right. \\
&+ (-1 \times 0.1)^2(0.084)^2 \\
&\left.\left.\left. + 2 \times (-1)(-1 \times 0.1)(-0.792)(0.028)(0.084)\right)\right]\right\} \\
&\times \left\{\exp\left[1.57\left((-1)^2(0.028)^2 + (-1 \times 0.1)^2(0.084)^2\right.\right.\right. \\
&\left.\left. + 2 \times (-1)(-1 \times 0.1)(-0.792)(0.028)(0.084)\right)\right] \\
&\left. - 1\right\}^{1/2} \times 100 \\
= &\ 2.5
\end{aligned}$$

where $(s_{7,7})^{1/2}$ and $(s_{13,13})^{1/2}$ are the standard deviations of the estimated regression coefficients $\hat{\beta}_7$ and $\hat{\beta}_{13}$, respectively, and are available in Table 2.

The Poisson regression model introduced in this paper can be developed and tested for other states in a similar manner. For those states in which detailed rural Interstate roadway and accident data are not available for conducting such an analysis, it is recommended that Model 5 be used with a slight modification, as follows:

$$\begin{aligned}
\hat{\mu}_i = \left(\frac{\text{AR}}{0.81}\right) v_i \exp(&-0.626471 + 0.02440x_{i,6} \\
&+ 0.088861x_{i,7} + 0.234209x_{i,13} + 0.077815x_{i,9} \\
&+ 0.033973x_{i,14} + 0.085763x_{i,11} - 0.025233x_{i,12}) \qquad (10)
\end{aligned}$$

where AR represents the overall truck accident involvement rate/MTM in recent years for the rural Interstate Highways in another state of interest, and 0.81 is the overall truck accident involvement rate/MTM for the road sections examined in this study. This modification is intended to adjust for the differences between Utah and the state of interest in, for example, weather and socioeconomic conditions, as well as the differences in accident reporting practices for nonfatal accidents and in the criteria used for classifying roadways. Under this modified model, the expected percentage reductions in truck accident involvements and associated standard deviations can still be computed from Equations 8 and 9 without any changes.

ACKNOWLEDGMENTS

This paper is based on the results of a research project sponsored by the Office of Safety and Traffic Operations Research and Development, Federal Highway Administration, the U.S. Department of Transportation. The opinions expressed in this paper are, however, solely those of the authors.

REFERENCES

1. *Fatal and Injury Rates on Public Roads in the United States.* FHWA-SA-90-029, Federal Highway Administration, U.S. Department of Transportation, June 1990.
2. L. Evans. *Traffic Safety and the Driver.* Van Nostrand Reinhold, New York, N.Y., 1991.
3. Cost and Safety Effectiveness of Highway Design Elements. In *National Cooperative Highway Research Program, Report 197,* Washington, D.C., 1978.
4. K. Rumar. The Role of Perceptual and Cognitive Filters in Observed Behavior. *Human Behavior and Traffic Safety* (L. Evans and R. C. Schwing, eds.) Plenum Press, New York, N.Y., 1985, pp. 151–165.
5. P. P. Jovanis and H. L. Chang. Modeling the Relationship of Accidents to Miles Traveled. In *Transportation Research Record 1068,* TRB, National Research Council, Washington, D.C., 1986, pp. 42–51.
6. S.-P. Miaou and H. Lum. Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accident Analysis and Prevention* (in press).
7. S. C. Joshua and N. J. Garber. Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models. *Transportation Planning and Technology,* Vol. 15, 1990, pp. 41–58.
8. S.-P. Miaou et al. *Development of Relationships Between Truck Accidents and Highway Geometric Design: Phase I.* Technical Memorandum, Oak Ridge National Laboratory, March and Nov. 1991.
9. F. F. Saccomanno and C. Buyco. Generalized Loglinear Models of Truck Accident Rates. Paper presented at Transportation Research Board 67th Annual Meeting, Washington, D.C., January 1988.
10. F. M. Council and J. F. Paniati. The Highway Safety Information System. *Public Roads,* Vol. 54, No. 3, Dec. 1990.
11. D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events.* Chapman and Hall, London, 1966.

12. A. C. Cameron and P. K. Trivedi. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics*, Vol. 1, 1986, pp. 29–53.
13. J. S. Cramer. *Econometric Applications of Maximum Likelihood Methods*. Cambridge University Press, New York, N.Y., 1986.
14. D. R. Cox. Some Remarks on Overdispersion. *Biometrika*, 70(1), 1983, pp. 269–274.
15. C. Dean and J. F. Lawless. Tests for Detecting Overdispersion in Poisson Regression Models. *Journal of the American Statistical Association*, Vol. 84, No. 406, 1989, pp. 467–472.
16. A. C. Cameron and P. K. Trivedi. Regression-Based Tests for Overdispersion in the Poisson Model. *Journal of Econometrics*, Vol. 46, 1990, pp. 347–364.
17. R. W. M. Wedderburn. Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, 54, 1974, pp. 439–447.
18. A. Agresti. *Categorical Data Analysis*. John Wiley, New York, N.Y., 1990.
19. B. W. Lindgren. *Statistical Theory*. Third Ed. Macmillian, New York, N.Y., 1976.
20. *A Policy on Geometric Design of Highways and Streets*. American Association of State Highway and Transportation Officials, Washington, D.C., 1990.
21. C. V. Zegeer et al. *Cost-Effective Geometric Improvements for Safety Upgrading of Horizontal Curves*. University of North Carolina, Chapel Hill, May 1990.
22. H. Bozdogan. Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions. *Psychometrika*, 52(3), 1987, pp. 345–370.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Comparison of Office of Motor Carriers Accident Data with Independent Truck Accident Data from Washington State

HOWARD S. STEIN

Detailed in this paper are the results of a comparison of accident data reported by motor carriers via the 50-T Form to the Office of Motor Carriers (OMC) with another independent source of similar accident data. This independent data source contains information from truck crashes that occurred on Interstate highways in Washington State that was collected as part of an independent truck safety study conducted by the Insurance Institute for Highway Safety (IIHS). The trucks in the independent data set were screened to determine which accidents should have been reported to OMC. Many comparisons were conducted to analyze the differences between trucks that reported to OMC versus those that did not and the accuracy of the information reported. This comparison found that only 40 percent of trucks involved in eligible crashes had their accident reported to OMC. Furthermore, many of the most important variables were not reported accurately. For example, of the 47 trucks with serious equipment defect identified in the Washington State Truck Study, only 3 reported defective equipment to OMC. Also, the truck configuration was reported incorrectly for about 20 percent of tractor-trailer trucks. Consequently, many past studies that have used the OMC 50-T form data for detailed analysis of truck safety may be invalid. On the basis of the results of these comparisons, several recommendations are made in this study to revise the 50-T form data and review its potential role in accident analysis.

Detailed in this paper are the results of a comparison of accident data reported by motor carriers via the 50-T Form to the OMC with another independent source of similar accident data. This independent data source contains information from truck crashes that occurred on Interstate highways in Washington State that was collected as part of an independent truck safety study conducted by the IIHS. The results of this Washington State Truck Study have already been reported in several journal articles (*1–3*). In addition to the standard police report form, the Washington State Truck Study also collected data via a supplementary truck form that was similar to the 50-T Form. Furthermore, an equipment inspection was performed for each crash-involved truck by commercial vehicle enforcement officers of the Washington State Patrol.

All the truck and police report data from the Washington State Truck Study were analyzed to determine which trucks should have reported their accidents to OMC. Afterward, an attempt was made to match all eligible crash-involved trucks from the Washington State Truck Study to trucks in the OMC accident files. The information reported by the commercial vehicle enforcement officers in the Washington State Truck Study was then compared with similar data reported to OMC by motor carriers on their 50-T Forms.

Contingency table analyses were conducted to investigate the agreement between the two data sources for truck characteristics (e.g., configuration, weight, and length), hours of service, motor carrier operation (e.g., type of carrier and fleet size), and crash circumstances (e.g., injury and property damage only and action of the truck). From these comparisons, recommendations are made for each data item studied concerning any bias in the OMC accident file that may affect carriers reporting to OMC and the accuracy of their information.

Kittelson & Associates, Inc., 610 SW Alder, Suite 700, Portland, Oreg. 97205.

WORK PLAN

This study was done as part of a OMC contract research project that reviewed the current use of information from the OMC 50-T Accident Form and its future status. OMC defined the scope and types of analyses that were to be conducted. OMC's primary objective was to examine whether there were any consistent patterns in the accuracy and completeness of the data that motor carriers were reporting to OMC via the 50-T Form. Specifically, OMC was concerned about how reporting of information varied by motor carrier characteristics, such as type of operation (i.e., common, contract, or private) and fleet size. This information was essential to OMC because the 50-T Form information was being evaluated to see how it was used by the government and other organizations, and whether reporting procedures or variables should be modified or vary by carrier characteristics.

In meetings with OMC, specific research questions defined by this comparison addressed

1. Are motor carriers accurately reporting vehicle defects?
2. How accurately are motor carriers reporting vehicle characteristics?
3. Are motor carriers accurately reporting hours of service?
4. What was the noncollision action?
5. Are injuries-fatalities accurately reported?

INDEPENDENT DATA SOURCES

Described in this section are the different sources of information that were used as comparison with the OMC accident data records. An example of each of the data sources is included in the complete report submitted to OMC (*4*). The role of each data source and its key variables used to match the other data files and the OMC accident records are noted in the following paragraphs. Also presented in this section are the variables that were used in the contingency table comparisons.

Washington State Patrol Accident Reports

The official Washington State Patrol accident report records were obtained from the Washington State Patrol on computer tapes. These computer files were processed to identify all truck accidents that occurred on Interstates or divided highways from 1984 to 1986. This process created records for 5,725 crash-involved trucks. This number is very large as it included both intrastate and Interstate carriers, all divided highways (not just Interstates), and the complete years from 1984 to 1986. As noted later, the truck crash data used in this comparison is for Interstate carriers only, crashes that occurred on Interstate system highways, and from June 1984 through June 1986. The key variables that were used to match other data files from the Washington State Truck Study were the date of the crash (year, month, day), time of the crash, Interstate, milepost, and the age of the driver. The information used from these records to compare with 50-T Form data were the injury and property damage estimates (to determine if the motor carrier was required to report to OMC), actions of the truck, and crash circumstances (e.g., contributing factors, driver citations).

Washington State Truck Study Supplementary Truck Forms

These forms indicated the same data used in the Washington State Truck Study, conducted from March 1984 through July 1986, that examined the role of truck, driver, and trucking operation characteristics in contributing to accident causation (*1–3*). They contained similar information to the OMC 50-T Form as well as additional information about the motor carrier fleet and driver characteristics. These supplementary forms were reviewed and all the major variables were entered into a computer file. The key variables used to match the other files were the date of the crash, time, Interstate, milepost, and the age of the driver. The variables used in the evaluation of the 50-T Form data were truck configuration, truck weight and length, truck operations, fleet size, driver experience, and hours of driving. Approximately 500 eligible trucks were identified.

Commercial Vehicle Inspection and Critical Item Inspection Forms

These forms recorded the results of the truck equipment inspections performed by the commercial vehicle enforcement officers on the crash-involved trucks of the Washington State Truck Study. These inspections are similar to the equipment inspections performed in the OMC Motor Carrier Safety Assistance Program (MCSAP). In fact, many of the MCSAP teams were trained by Washington State personnel. These forms indicated the condition of the truck's major component systems such as brakes, steering, and tires. The officers completed a critical item form indicating whether any of these systems were defective and if these defects constituted a violation of operating requirements or, if more serious, required that the truck be placed out of service. For example, one brake out of adjustment would constitute a violation, but having 25 percent of the brakes out of adjustment would require that the truck be immediately placed out of service until repairs are completed.

The driver's log book status was also noted on these inspection forms. A log book violation consists of having an incorrect log that is less than 24 hrs behind. An out-of-service log book violation consists of having a log book violation that is more than 24 hrs behind, violation of the driver hours of service rules, or operating without a log book.

This file contained fewer trucks than the supplementary form file because in some instances the crash-involved truck was severely damaged in the crash or circumstances did not permit inspections to be conducted, such as in a blizzard. The key variables in this file were the same as the supplementary form file. The results of these inspections were compared with the condition of the truck and mechanical defects that should have been noted by motor carriers on 50-T Forms.

All of these files (State Patrol reports, supplementary truck forms, and vehicle inspections) were merged together using the key variables as noted. There was little trouble matching the Washington Truck Study data together, but there were some difficulties in matching this data with the police reports. This occurred primarily because of rounding of key data items, in particular, the time, milepost, and driver age. A straightforward computer match on the key variables resulted in only about 50 percent of the trucks matching. Consequently, each record in the police file was carefully reviewed to determine whether it matched with a truck record from the Washington Truck Study. In almost all cases, the problem was that one of the key variables in the police file was slightly different from the same variable recorded in the Washington Truck Study. For example, the milepost may be recorded as 269 on the police report, but 270 on the Truck Study forms, or the time was listed as 800 on the police report but 755 on the Truck Study forms.

MATCHING WASHINGTON TRUCK STUDY DATA WITH OMC FILES

The next step was to match trucks from the Washington Truck Study with truck accident reports in the OMC 50-T Accident Form files. Computer files of the 50-T Form data were provided by OMC. The OMC files were first screened to identify truck crashes that occurred on Interstates during the period 1984 to 1986 in Washington State. This resulted in 468 OMC truck accident reports. The key variables in this file that were used to match crashes in the Washington State data were the date of the crash, Interstate route, time, and driver age. The limitation in this matching process was the few variables in the OMC data that were available to pin down to location and circumstances of the crash that were not variables that would be used later as part of our analysis. However, there were only a few instances of two Interstate trucks being involved in a crash together or on the same day on the same Interstate at approximately the same time, mostly during poor weather. Consequently, these "matching" variables almost always defined a unique event. Driver age variables were used to help match up these accident records because it was felt that the driver's age would be reported accurately.

An initial attempt to match the two data sources revealed that there were also problems with the rounding off of key variables. Consequently, the OMC data were reviewed by hand for better matching. This review found that in the OMC file there were many instances in which the time was not given in military time and the age was transposed (i.e., 32 rather than 23). In almost all cases in which a truck was present in the Washington State data, but not in the OMC file, there was no truck accident report in the OMC file for that date, on that Interstate, or within many hours of the appropriate time. None of the other study variables were used for matching as they were to be part of the analysis. This review identified 185 trucks that matched between the Washington State data and the OMC files. The remaining OMC trucks were simply not captured as part of the Washington State study. The truck inspectors in the Washington State Truck Study did attempt to go to most major truck crashes during the study period, but they could not investigate all crashes. For example, in urban areas there may have been more than one crash occurring at a given time, precluding investigation of both crashes. In addition, in the more rural areas there may not have been a truck inspector available to go to the crash site in a timely manner, particularly if the truck could drive away from the crash site.

The remaining eligible trucks in the Washington State data were then screened to eliminate truck crashes that may not have met the OMC reporting criteria of the accident resulting in an injury where treatment was received away from the site or meeting the property damage reporting criteria (at least $2,000 in 1984 and 1985 and $4,200 in 1986). This screen identified 287 trucks that were in the Washington State Truck study that should have been reported to OMC, but were not present in their files. Consequently, the final analysis file contains 472 truck records; 185 matched with OMC reports and 287 that were not matched.

The remaining trucks in accidents from the OMC files were dropped from further analysis. Given the poor accuracy of many of the key variables reported to OMC that is documented later in this paper, performing additional analysis of these trucks would not be valid.

DATA ANALYSIS

The data analysis consisted of generating two-way (two-factor) contingency tables that classified the trucks by various characteristics and computing the Chi-squared statistic to determine whether the factors being compared are independent. (In most comparisons, the cell sample sizes were too small for this statistic to be reliable.) Because more than 100 variables were available for analysis, a specific work plan was submitted to and approved by OMC, as detailed earlier in this paper. The specific research questions identified by OMC addressed aspects of defective truck equipment, vehicle characteristics, hours of service, driver condition, collision events, crash outcome (property damage/injury/fatal) and driver experience. Where appropriate, these comparisons were also performed to include a third factor such as carrier type (private, contract or common) or fleet size. The analysis was conducted using the Statgraphics (Version 2.6) statistical software. For ease of presentation, only summary tables are presented in this paper, as well as several sample comparison contingency tables. All the detailed analyses were included in appendices of the main report submitted to OMC (*4*).

RESULTS OF COMPARING OMC DATA WITH WASHINGTON STATE TRUCK STUDY

Detailed in this section are the results of analyses conducted to address the specific research questions that were identified by OMC, as well as a short description of the importance of the research question and the implication of our findings.

Completeness of the OMC File

Other studies have reported that only 40 percent of trucks in accidents eligible for the OMC files are actually reported, but there has been little documentation of this or what characteristics affect this bias (*5,6*). This study found that 39 percent of the trucks involved in OMC-eligible crashes that occurred on Interstate highways in Washington State during the study period reported to OMC (185 out of 472). There is little reason to believe that this finding is peculiar to Washington State because all of these trucks were Interstate carriers that operate throughout the United States. If anything, this finding might be considered conservative. It could be argued that because these truck drivers knew they were becoming part of a Washington State Patrol "study," carriers might have an incentive to report their accidents because of the special attention and possible follow up that might occur.

The issue of reporting (or matching) was also investigated by several other factors. Classified by type of carrier operation, common carriers reported more frequently (46 percent) compared with contract (31 percent) or private (24 percent) carriers (see Figure 1). Large fleets, containing more than 50 trucks, reported better (58 percent) than medium-size fleets containing 11 to 50 trucks (36 percent) or small fleets with 10 or fewer trucks (24 percent) (see Figure 2). Both of these trends were statistically significant. Trucks without defective equipment reported more frequently (42 percent) than trucks with out-of-service defects (34 percent). Our analysis also found that crashes involving younger drivers (30 years old or younger) tended to report less (33 percent) compared with older drivers (41 percent). Finally, little difference was found in reporting for tractor trailer trucks versus doubles (both about 40 percent), but crashes involving single-unit trucks were reported (24 percent) less frequently.

Are Motor Carriers Accurately Reporting Equipment Defects?

The issue of reporting defective equipment is of critical concern to OMC because defective equipment is key to assigning

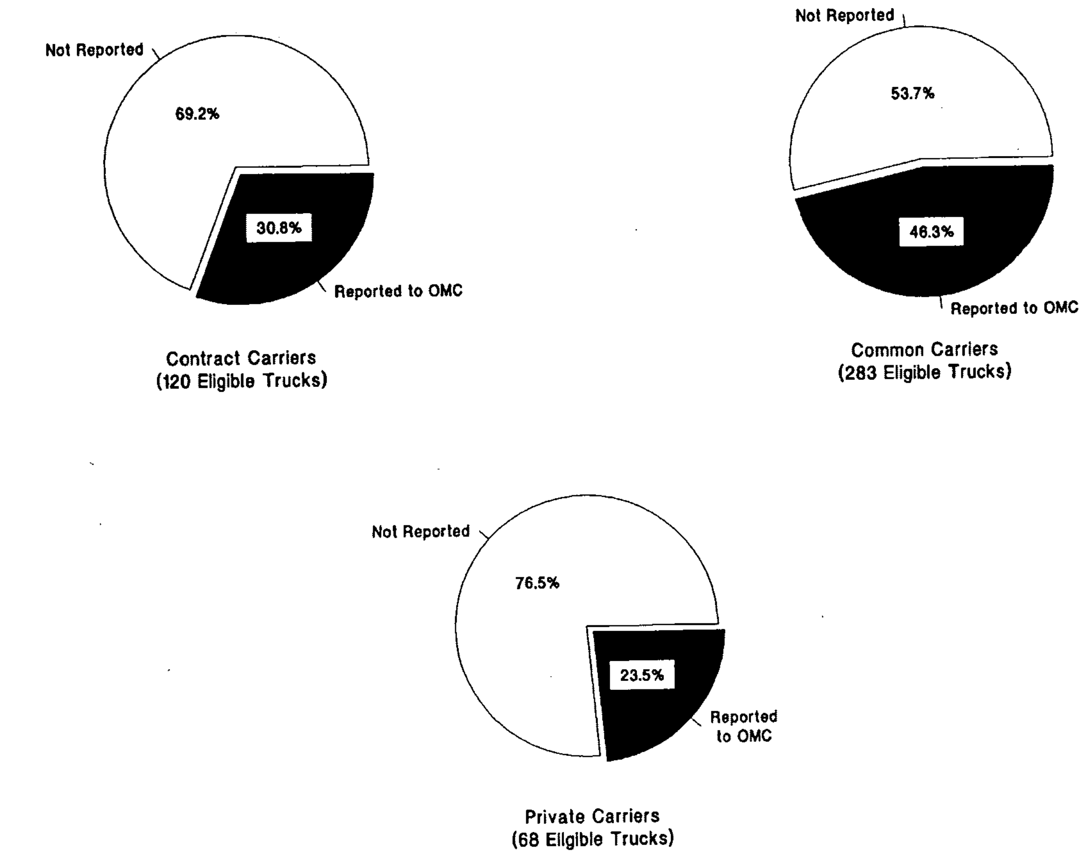


FIGURE 1 Reporting of crashes to OMC by carrier type.

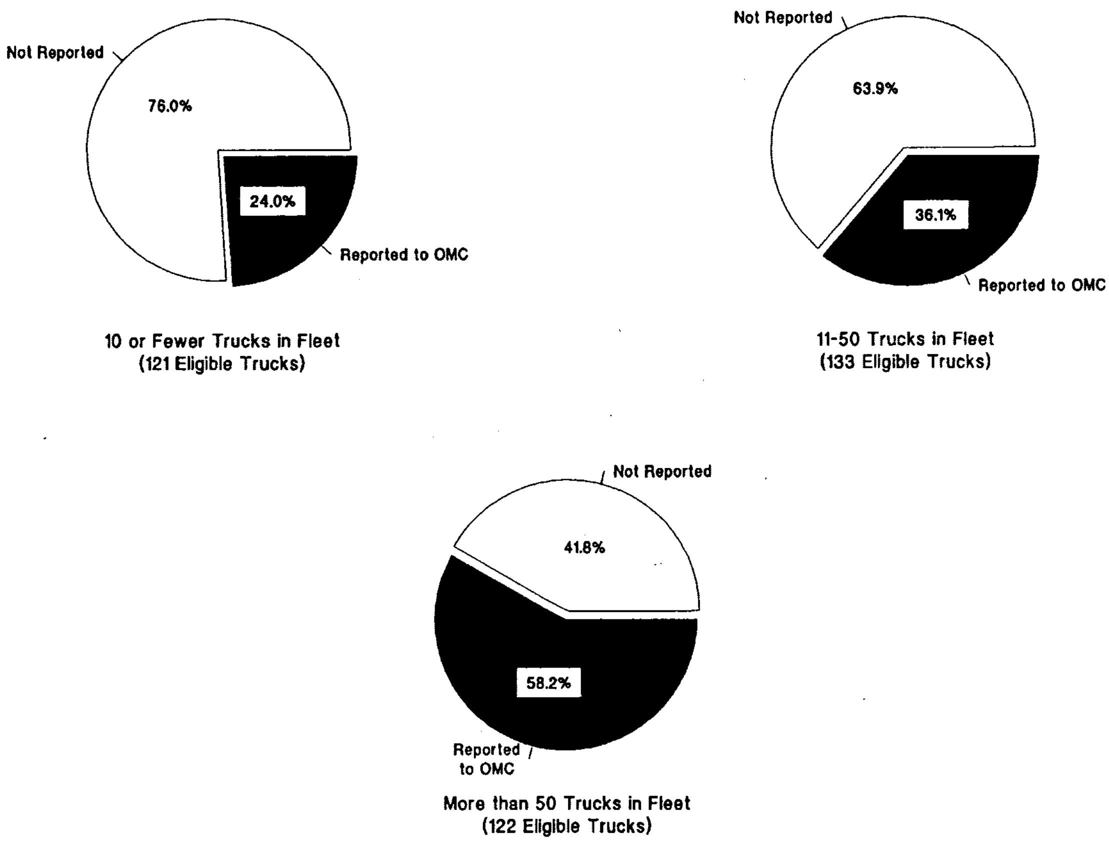


FIGURE 2 Reporting of crashes to OMC by fleet size.

preventability to truck crashes. In the past, OMC data has indicated that only 5 percent of trucks in crashes have defective equipment. At the same time, random roadside inspections of trucks indicate that as many as 50 percent of trucks have defective equipment (*5,6*). Although all the equipment defects may not be related to specific crash circumstances simply based on random occurrence, it would be expected that the proportion of trucks in the OMC data should be significantly higher than 5 percent.

Comparisons of the equipment defect variable in the OMC 50-T Form data with the results of the equipment inspections performed by the Washington State Patrol indicated that truck equipment defects are rarely reported to OMC. As indicated in Table 1, of the 47 trucks identified by the Washington State Patrol as having out-of-service defects, only 3 (6 percent) reported having defective equipment to OMC. As mentioned previously, having defective equipment may not be the most critical factor in all these accidents, but most of these defects involved brakes or steering and it would be difficult to determine crashes in which braking and steering are not relevant. Of the 47 trucks with out-of service defects, 66 percent (31 of 47) had out-of-service brake adjustment defects. In addition, 10 percent (5 of 47) had (separately or in combination) out-of-service steering defects; typically, too much play in the steering wheel without any response.

Also noted in our analyses, trucks with out-of-service defects tended to report less frequently (31 percent) to OMC than trucks with no defects or where an inspection could not be completed (41 percent). This poor reporting of trucks with serious defects to OMC may reflect the reluctance of carriers to submit 50-T Forms for an accident where they may be at fault.

One problem with this comparison is the actual question asked on the OMC form: Were mechanical defects or failures apparent on your vehicle at the time of the accident? This question is not specific and can be misinterpreted. For example, how is apparent defined? Although the OMC instructions go into this issue in detail (requiring that each defect known to exist before the accident, brought to light by the accident, or discovered by investigation of the accident should be recorded), few carriers probably review these additional instructions. A more direct question would be: Were any equipment defects present on your vehicle at the time of the accident? On the other hand, it would be difficult to believe that a substantial portion of the drivers with trucks having out-of-service brake and steering defects are not aware of the problems.

Despite the problem described, of the 185 matched trucks, private carriers reported defects to OMC more accurately (75 percent) compared with either common or contract carriers (both about 45 percent). Medium-size fleets (11 to 50 trucks) reported defects less accurately (35 percent) to OMC than either of the other two fleet categories (both more than 50 percent).

Additional analyses were conducted to determine how defective equipment varied among the motor carrier operation variables for all 472 trucks involved in the Washington State study. Contract carriers had the highest proportion (37 percent) of trucks with out-of-service defects compared with private (33 percent) or common carriers (27 percent) (Figure 3). Smaller fleets (10 trucks or less) had a higher proportion of trucks with out-of-service defects (38 percent) compared with larger fleets (28 percent) (Figure 4). Finally, it was found that tractor trailer trucks (33 percent) had out-of-service defects more frequently than doubles (20 percent) or single unit trucks (23 percent).

Are Motor Carriers Accurately Reporting Hours of Service?

Truck drivers are required to maintain an accurate log of their activities, specifically the number of hours of driving and rest they have had while on and off duty. There is significant

TABLE 1 Reporting of Defective Equipment to OMC

DEFECTIVE EQUIPMENT NOTED IN WASHINGTON STATE TRUCK STUDY:	CONDITION OF TRUCK EQUIPMENT REPORTED TO OMC: No Defects	Defects	Row Total
No Defects	79 (96.3)*	3 (3.7)	82 (45.8)
Equipment Violations	49 (98.0)	1 (2.0)	50 (27.9)
Out-of-Service Violations	44 (93.6)	3 (6.4)	47 (26.3)
Column Total	172 (96.1)	7 (3.9)	179** (100.0)

* Numbers in () are percents. Cell percents are by row.
** Table does not total to 185 because truck inspections were not performed by Washington State Patrol for 6 trucks due to adverse conditions.

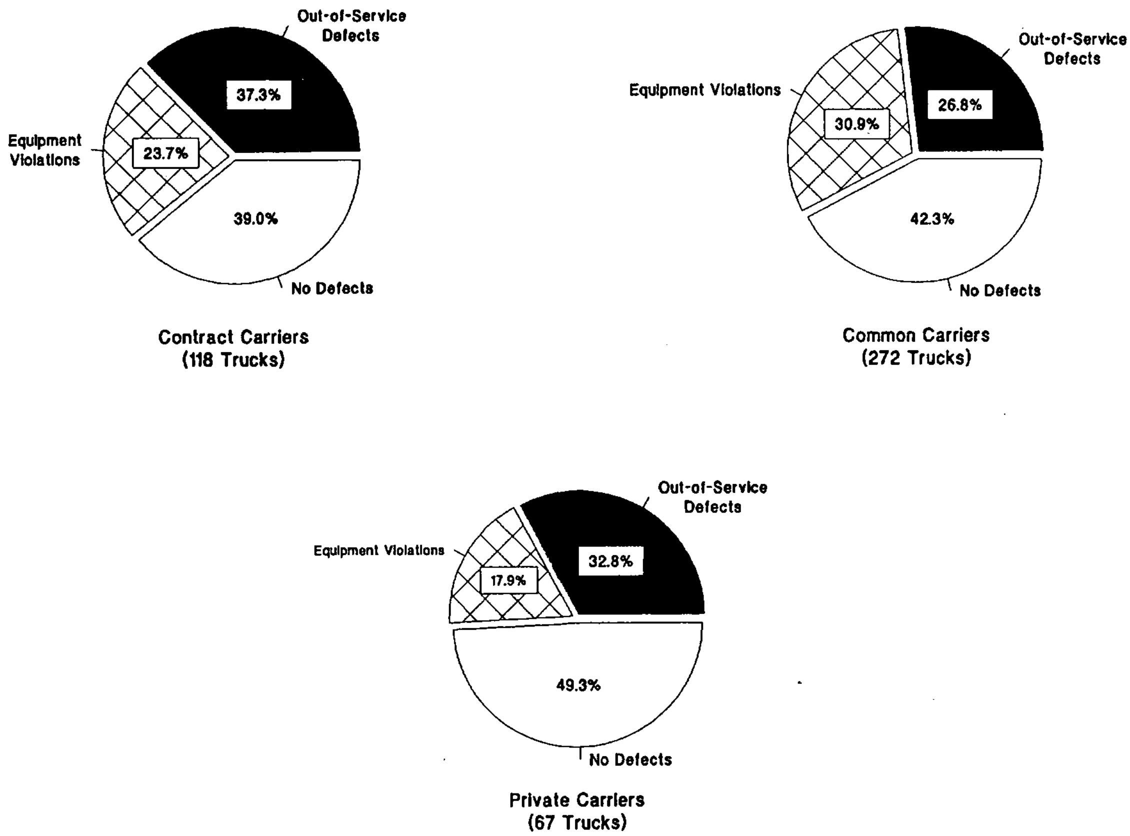


FIGURE 3 Incidence of defective equipment for crash-involved trucks in the Washington State Truck Study by carrier type.

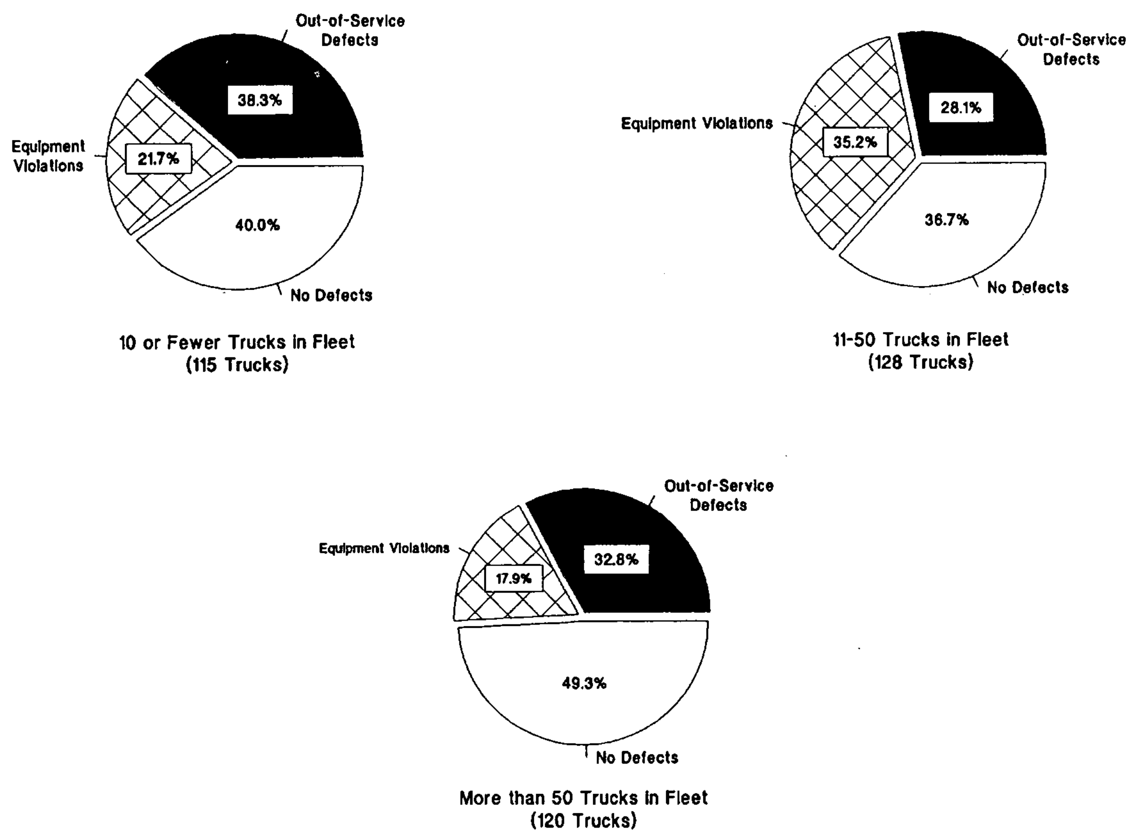


FIGURE 4 Incidence of defective equipment for crash-involved trucks in the Washington State Truck Study by fleet size.

evidence that drivers routinely exceed these hours of service rules and that these log books are not properly maintained. There also appears to be growing support for automatic onboard recording devices to monitor driver hours rather than continuing to use manual log books. On the other hand, some accident studies have found that the majority of truck crashes occur during the first several hours of the trip and, thus, excess hours of service may not be as large a factor in accident causation as believed.

Our analyses of the OMC hours-driven variable versus the hours of service recorded in the Washington State Truck Study indicated that the hours of service category reported to OMC is generally accurate. Almost 50 percent of the "matched" truck drivers had been on the road 3 hrs or less, but 5 percent had been on the road for 10 hrs or more. However, these results are biased by the fact that they do not include many of the drivers who had deficiencies in their log books. Comparison is made in Figure 5 of the log book status of those drivers that matched versus those that did not, and indicates that 60 percent of the drivers with log book violations (including out-of service) were not found in the OMC data. In addition, our analysis found that contract carriers reported their hours of driving (within 2 hrs of the Truck Study data) less accurately (73 percent) compared with the other carriers (both about 85 percent). Drivers in smaller fleets tended to report their hours of driving more accurately (89 percent) than drivers in larger fleets (both about 80 percent).

The frequency of log book violations among all trucks in the Washington State Truck Study was also examined by other motor carrier operation variables. Driver out-of-service log book violations were about twice as frequent among contract carriers (16 percent) than for either common or private carriers. Log book out-of-service violations were slightly more frequent among small fleets (14 percent) than medium size or large fleets (9 percent). Finally, 13 percent of tractor trailer drivers in crashes had out-of-service violations compared with only 5 percent of drivers of double trailer trucks.

What Was the Noncollision Event?

For single-vehicle truck accidents, the OMC 50-T Form asks carriers to report what other one noncollision event occurred, including ran-off-road, jackknife, overturn, and fire. The implication is that some action of the driver caused the crash and that it could have been avoided perhaps by going slower, the driver paying more attention to the roadway, or having the driver adjust his behavior for adverse driving conditions. The same data were also collected as part of the Washington State Truck Study except that more than one event could be reported.

Sixty-four of the 185 matched trucks reported their crashes as noncollision crashes, 81 were reported as collisions involving another moving vehicle, and 39 were coded as collisions

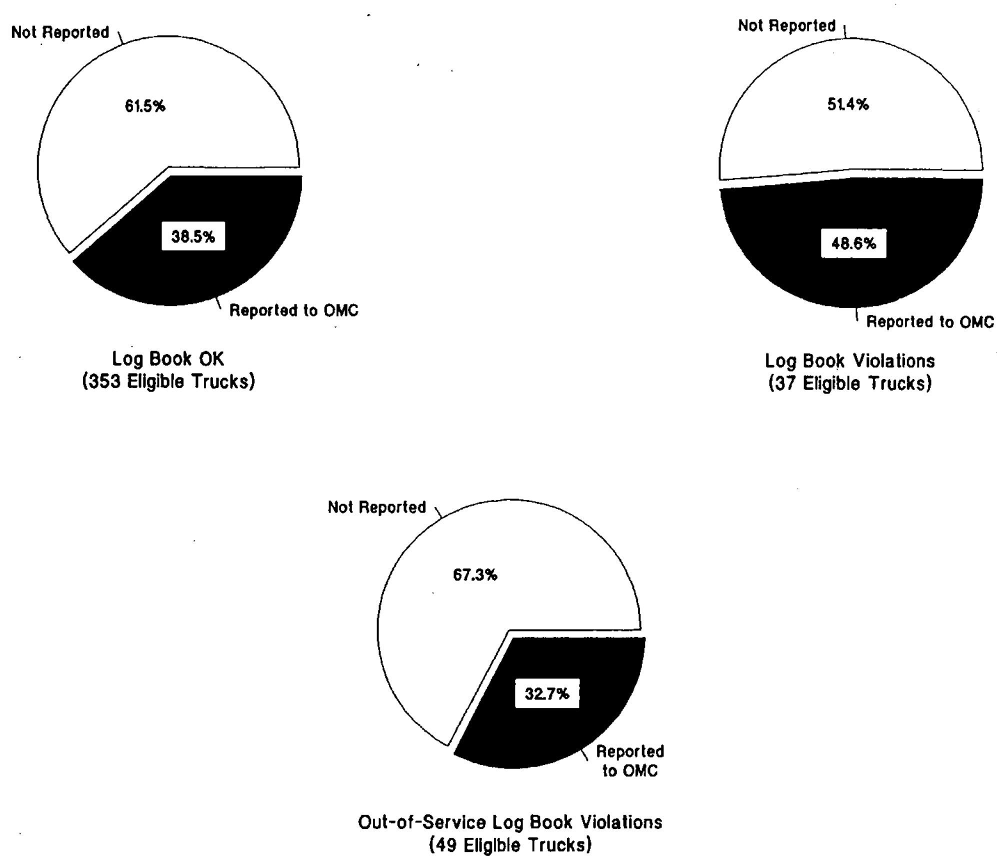


FIGURE 5 Reporting of crashes to OMC by status of log book.

with fixed objects. These types of collisions varied by carrier type, but were similar among the fleet size categories. Private carriers had proportionally fewer noncollision crashes and more multivehicle crashes than the other carrier types.

Comparing the noncollision events reported to OMC with similar events recorded in the Washington State Truck Study indicated that there is good agreement between the two data sources. When an event was reported to OMC, it was also recorded by the Truck Study. However, the problem is that during a single vehicle noncollision accident, several of these events could occur. For example, a truck can jackknife, run off the road, and then overturn down the side of the road. The OMC data records only one event. Consequently, this OMC variable may not tell the complete story of single-vehicle crash events. For example, of the 64 matched trucks coded as noncollision in the OMC file, the Washington State Truck Study recorded that 42 of them overturned (see Table 2). However, only 25 of these trucks were reported to OMC as overturning, whereas 12 were reported as simply run off the road, and 4 were reported as jackknife crashes. If this OMC variable is to be useful, all applicable events need to be recorded. In addition, these crash events can also occur as part of either multivehicle or fixed object crashes, and, consequently, there is no reason that crash events are not reported for these crash types.

How Accurately Are Vehicle Characteristics Reported?

Comparing truck length between the two data sources indicated that almost 75 percent of the truck lengths reported to OMC were within 5 ft of those recorded in the Washington State Truck Study. These results did not differ significantly by truck type or fleet size. However, the accuracy of reported truck length did differ by carrier type with common (78 percent) and contract (69 percent) carriers being more accurate than private carriers (57 percent).

The differences between the truck weight reported to OMC and recorded in the Washington Truck Study were not in as close agreement as truck length. Overall, only 55 percent of the matched trucks reported their weight within 5,000 lbs of the weight recorded in the Washington State Truck Study. In contrast to the truck length results, the differences in reported weight did not differ significantly by carrier type, but smaller fleets (10 or fewer trucks) reported their weight more accurately (67 percent within 5,000 lbs) than larger fleets (52 percent within 5,000 lbs).

Perhaps the most controversial truck variable that has been examined by many truck studies is truck configuration (*1*,*5*,*6*). In particular, most truck safety studies have compared the accident record of tractor trailer trucks with double-trailer configuration trucks (tractor with two trailers). Unfortunately, the results of this data comparison study indicate that many of these past accident studies may contain serious errors. Reconciling the truck configuration variables between the two data sources analyzed in this study indicate that only 75 percent of the matched trucks were classified as the same configuration in both the OMC and Washington State Truck Study data (see Table 3). Only 82 percent of tractor trailers were classified as this configuration in the OMC data; 16 percent were classified as doubles. Similarly, only 77 percent of doubles were classified as this configuration in the OMC

TABLE 2 Comparison of Crash Events Between Washington State Truck Study and OMC 50-T Form

Sole Crash Event Reported to OMC on the 50-T Form:	Truck Overturn was **One of the Crash Events** Reported By Washington State Patrol: NO	YES	Row Total
Ran-Off-Road	5 (22.7)*	12 (28.6)	17 (26.6)
Jackknife	16 (72.7)	4 (9.5)	20 (31.3)
Truck Overturn	0 (0.0)	25 (59.5)	25 (39.1)
Truck Units Separated	1 (4.5)	0 (0.0)	1 (1.6)
Other	0 (0.0)	1 (2.4)	1 (1.6)
Column Total	22 (34.4)	42 (65.6)	64 (100.0)

Note: Only one crash event is recorded on the OMC 50-T form. In contrast, the Truck Study recorded as many events as apply to the crash.

* Numbers in () are percents. Cell percents are by column.

TABLE 3 Comparison of Truck Configuration Between Washington State Truck Study and OMC 50-T Form

	Truck Configuration Recorded by Washington State Truck Inspector					
Truck Configuration Reported on 50-T Form to OMC	Tractor	Tractor Trailer	Tractor 2 Trailers	Single Unit Truck	Truck-Trailer	Row Total
Truck (A)*	1 (20.0)*	0	1 (20.0)	1 (9.1)	0	3
Truck Trailer (AD)	0	1 (0.7)	0	4 (36.4)	4 (36.4)	9
Truck-Other (AF)	0	0	0	0	1 (14.3)	1
Tractor (B)	2 (40.0)	1 (0.7)	1 (3.8)	0	0	4
Tractor Trailer (BC)	1 (20.0)	111 (82.2)	3 (11.5)	4 (36.4)	2 (28.6)	121
Tractor 2 Trailers (BCD)	1 (20.0)	22 (16.3)	20 (76.9)	1 (9.1)	0	44
Triple (BCDF)	0	0	1 (3.8)	0	0	1
Tractor-Other (BF)	0	0	0	1 (9.1)	0	1
Column Total	5 (100.0)	135 (100.0)	26 (100.0)	11 (100.0)	7 (100.0)	

* Truck Unit codes from OMC 50-T Form.
* Numbers in () are percents. Cell percents are by column

data; 12 percent were classified as regular tractor trailer trucks. Furthermore, more than half the doubles (24 of 44) reported to OMC were actually some other configuration. Consequently, these analyses indicate that not only were there underreporting differences to OMC by truck type, but at least 1 out of 4 trucks had their configuration reported incorrectly. Contract carriers reported their truck configurations more accurately (84 percent) than either common (71 percent) or private (75 percent) carriers. Also, larger fleets reported their truck configurations more accurately (78 percent) than smaller fleets (69 percent).

As a general observation, the analyses discussed reveal that for some data, such as truck length, there seems to be general agreement among the two data sources, but for the other data, such as truck weight and configuration, there were serious differences between these two data sources.

Are Injuries and Fatalities Accurately Reported?

Overall, 38 percent of all matched crashes involved property damage only, 59 percent had someone injured, and 3 percent involved a fatal injury. All the crashes in the matched data file that resulted in someone being fatally injured involved common carriers. The distribution of crashes that were reported to OMC tended to be more severe than the general sample of crashes, particularly for contract and private carriers and medium-size fleets (11 to 50 trucks). This reinforces the theory that carriers would tend to better report crashes that might be investigated further or in depth. There was some discrepancy about the number of persons injured that may have arisen because of minor injuries and how they are categorized. Also, truck drivers typically leave the crash scene once basic information has been collected and the police have everything under control. Consequently, truck drivers would not follow up on the actual total number of people injured or receiving treatment afterwards.

DISCUSSION AND CONCLUSIONS

Compared in this study have been truck accident data collected as part of an independent safety study with similar data that was self-reported to OMC by motor carriers via their 50-T Form. Overall, it was found that only about 40 percent of eligible crash-involved trucks reported their accidents to OMC. This finding is consistent with the results of other studies. The lack of reporting varied by several factors, such as truck type and fleet size, that could have significantly affected the results of previous safety studies that used the OMC 50-T Form data as their basis for compiling accident frequencies, rates, or their characteristics.

As a general trend, of the data items compared between the two data sources, there was good agreement among the same data items such as truck length. However, there were some serious deficiencies in the agreement of the other data such as defective equipment and truck configuration. By far the biggest deficiency was that defective truck equipment was rarely reported to OMC, even though more than 25 percent of trucks had out-of-service equipment defects.

Summarized by carrier type and fleet size in Tables 4 and 5 are the major results of this study. Contract and private carriers reported their crashes to OMC less frequently than common carriers. Of the data reported to OMC, there was no clear pattern of reporting bias among the carrier types. However, many of the poor accident reporting and crash factor characteristics were associated with contract carriers, such as having a high percent of noncollision crashes, and worse reporting of hours of service and defective equipment. Considering fleet size, there was a clear pattern; the larger the fleet, the more frequently the crash was reported to OMC. Of the data reported to OMC, many of the more serious data deficiencies were associated with medium-size fleets (11 to 50 trucks), such as inaccurate reporting of defective equipment and truck weight, and having a high percentage of crashes involving injuries.

The implications of these findings depend on what is expected of the OMC data file and the accuracy desired for the various data elements. If the OMC file is simply to document the numbers of crashes that occur and some basic characteristics of truck crashes, the 50-T Form and reporting procedures might be adjusted for the deficiencies noted in this study. Given this scenario, OMC should consider dropping the more controversial data items, specifically defective equipment. A follow-up study should then be conducted to determine if more carriers were reporting their accidents, thereby enhancing the completeness of the OMC file without requiring significant efforts to monitor carrier compliance with the 50-T Form requirements.

TABLE 4 Summary of Findings from OMC–Washington State Truck Study Comparisons by Type of Carrier

	Type of Carrier		
	Common	Contract	Private
Percent of Eligible Crashes in Washington State Truck Study that were Matched with 50-T Form File	46	31	24
Compared to Washington State Truck Study:			
Percent Accurately Reporting Truck Equipment Defects on 50-T Form (see Text)	43	45	75
Percent Reporting Hours Driving on 50-T Form (within 2 Hours)	84	73	86
Percent Reporting Non-Collision Crashes on 50-T Form	30	38	19
Percent Reporting Truck Length on 50-T Form (within 5 Feet)	78	69	57
Percent Reporting Truck Weight on 50-T Form (within 5000 lbs)	56	53	53
Percent Reporting Same Truck Configuration on 50-T Form	71	84	75

TABLE 5 Summary of Findings from OMC–Washington State Truck Study Comparisons by Fleet Size

	Fleet Size:		
	1-10	11-50	More than 50
Percent of Eligible Crashes in Washington State Truck Study that were Matched with 50-T Form File	24	36	58
Compared to Washington State Truck Study:			
Percent Accurately Reporting Truck Equipment Defects on 50-T Form (see Text)	54	35	51
Percent Reporting Hours Driving on 50-T Form (within 2 Hours)	89	79	82
Percent Reporting Non-Collision Crashes on 50-T Form	31	25	37
Percent Reporting Truck Length on 50-T Form (within 5 Feet)	71	72	78
Percent Reporting Truck Weight on 50-T Form (within 5000 lbs)	67	47	57
Percent Reporting Same Truck Configuration on 50-T Form	69	78	79

If the OMC file is to be used for more in-depth comparisons of truck safety (as it has been done in many past studies), its current deficiencies constitute significant biases that may cause these types of studies to be invalid. Specifically, key deficiencies include mechanical defects, truck configuration, and crash events. Few mechanical defects were reported to OMC, although they were commonly found in more than half the trucks inspected in the Washington State Truck Study. There were also significant discrepancies in the reporting of the truck's configuration that may affect the results of the many studies that used the OMC file to compare the accident rates of tractor trailers versus doubles. In addition, studies that used this OMC data to estimate the frequency of crash characteristics such as truck overturn or jackknife have significantly underestimated the occurrence of these events.

As an alternative, OMC could develop a system in which their current data reporting is supplemented by crash investigations using MCSAP truck inspectors similar to the ones conducted in the Washington State Truck Study. Given that a staff of trained truck inspectors already existed in Washington State (and does nationally as part of the MCSAP program), the average cost was less than $200 for each in-depth truck inspection conducted by the Commercial Vehicle Enforcement officers of the Washington State Patrol. To reduce costs to OMC, the crashes they would investigate could be limited to the more serious crashes or of special interest (such as the one in which the truck overturned), which would probably merit special investigation by the local police authorities anyway.

In either case, better enforcement and auditing of the 50-T Form requirements and carrier compliance are needed. In addition, to assist OMC in auditing of carriers and conducting analyses such as was performed in this project, the 50-T Form information should be directly linked to other accident documents by including proper identification information that would directly link the 50-T Form to local police reports or other similar information.

ACKNOWLEDGMENT

The work detailed in this paper was performed while the author was with Callow Associates of Reston, Virginia (which was recently acquired by Science Applications International Corporation), under subcontract to the Scientex Corporation as part of their contract with OMC entitled "Accident Reporting System Augmentation, Task E: Feasibility of Expanded SAFETYNET Accident Module." The author would like to thank Cary D. Vick and Daniel Gray with Callow Associates for their assistance in performing this work and preparing this article.

REFERENCES

1. H. S. Stein and I. S. Jones. Crash Involvement of Large Trucks by Configuration: A Case-Control Study. *American Journal of Public Health*, Vol. 78, No. 5, May 1988.
2. I. S. Jones and H. S. Stein. *Effect of Driver Hours of Service on Tractor-Trailer Crash Involvement*. Insurance Institute for Highway Safety, Washington, D.C., Aug. 1990.
3. I. S. Jones and H. S. Stein. Defective Equipment and Tractor-Trailer Crash Involvement. *Accident Analysis and Prevention*, Vol. 21, No.5, 1989.
4. H. S. Stein. *Comparison of OMC Accident Data with Independent Data from Washington State*. Callow Associates, Reston, Virginia, 1989.
5. *Big Trucks*. Insurance Institute for Highway Safety. Washington, D.C., 1985.
6. *Gearing Up for Safety: Motor Carrier Safety in a Competitive Environment*. Office of Technology Assessment, OTA-SET-382, Sept. 1988.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Truck Accident Models for Interstates and Two-Lane Rural Roads

YUSUF M. MOHAMEDSHAH, JEFFERY F. PANIATI, AND ANTOINE G. HOBEIKA

Truck travel increased from 668 billion km (400 billion vehicle mi) of travel to 1,002 billion vehicle km (600 billion vehicle mi) from 1980 to 1989, a staggering 50 percent increase. If this trend continues, truck travel will exceed 1.67 (1 trillion vehicle mi) trillion vehicle km by the end of the year 2000. This increase poses operational and safety problems for both passenger vehicles and trucks. To improve the safety of existing highway facilities and to determine the design standards for new truck facilities, an understanding of the relationship between truck accidents and the geometry of the highway is required. The objectives of this study were to identify the roadway variables that affect truck accidents and to develop mathematical models of their relationships. Data from the Highway Safety Information System were used in this analysis. The Highway Safety Information System is a new data base developed by the Federal Highway Administration. It contains accident, roadway, and traffic data from five states. Models for truck accidents on Interstates and two-lane rural roads were developed using data from the state of Utah. The Interstate model indicates that truck accidents are primarily affected by horizontal curvature and vertical gradient. For two-lane rural roads, the model indicates that truck accidents are affected by the shoulder width and the horizontal curvature. Gradient was not found to have an effect on truck accidents on two-lane roads, although this may be because of inadequate data.

The economy of the United States is largely based on freight transportation and most of this freight movement takes place through highways by means of trucks. Travel data show that truck travel increased from 668 billion vehicle km (400 billion vehicle mi) to 1,002 billion vehicle km (600 billion vehicle mi) from 1980 to 1989, a staggering 50 percent increase (*1*). If this trend continues, truck travel will exceed 1.67 trillion vehicle km (1 trillion vehicle mi) by the end of this century. This increase in truck travel causes a number of operational and safety problems on the highway. These problems result from the shear dimension of the trucks as well as their acceleration and deceleration characteristics.

To improve the safety of existing highways, a clear understanding of the relationship between truck accidents and the design of the highway is needed. To achieve this, a mathematical model of the relationship between truck accident rates and roadway design variables is required.

A number of models have been developed in the past. However, they are single variable models based on only one or two years of accident data and on a limited amount of roadway mileage. Hence their ability to explain truck accidents is limited. Documented in this paper is the development of a truck accident model for different highway types using a data base recently developed by the Federal Highway Administration (FHWA) called the Highway Safety Information System (HSIS).

Y. M. Mohamedshah, AEPCO, Inc., 6300 Georgetown Pike, McLean, Va. 22101. J. F. Paniati, Office of Safety and Traffic Operations Research and Development, Federal Highway Administration, 6300 Georgetown Pike, McLean, Va. 22101. A. G. Hobeika, University Center for Transportation Research, Virginia Polytechnic Institute and State University, 106 Faculty Street, Blacksburg, Va. 24061.

PREVIOUS RESEARCH

An extensive literature review was conducted to determine the causes of truck accidents, identify the critical variables affecting accidents in general and truck accidents in particular, and examine the accident models developed in the past.

Truck Characteristics

The occurrence of truck accidents is different from that of passenger vehicle accidents because of the special characteristics of trucks (*2*):

1. Trucks are much heavier and larger in dimension compared with passenger cars;
2. Trucks have less effective acceleration capabilities than passenger cars and have greater difficulty maintaining their speeds on upgrades; and
3. Trucks have a lower deceleration in response to braking than do passenger cars.

Because of these differences, trucks are affected differently by roadway characteristics, and truck accidents tend to be more severe than those involving passenger cars. Although studies of passenger car accidents can provide insights into important highway variables, a complete understanding of truck-highway relationships requires the use of truck accident data.

Critical Geometric Features

Several studies have examined the critical geometric features affecting truck accidents. In an FHWA report (*3*) on improving truck safety, six major design deficiencies for interchanges causing rollover and jackknifing truck accidents have been identified. Examples are (a) abrupt changes in compound curves, (b) short deceleration lanes on tight radius, and (c) steep downgrade at the exit ramp. In a report on hazardous material (HAZMAT) accidents by FHWA (*4*), several geometric features that cause truck accidents have been identified, such as (a) number of lanes, (b) lane width, (c) shoulder width, (d) median width, (e) alignment, (f) surface condition and (g) pavement condition.

Out of these geometric features, the most important in truck accident occurrence are vertical gradients and horizontal curves. Downgrades may lead to an excessive increase in truck speeds resulting in runaway accidents and rear ending of slow-moving vehicles, whereas on upgrades the truck moves slowly, resulting in rear ending of trucks by fast-moving vehicles. Horizontal curves can contribute to rollover problems for trucks with high centers of gravity or when the load shifts while negotiating the curve. On two-lane roads, trucks may encroach on the opposite lane while negotiating a curve, posing a hazard for opposing vehicles (C. V. Zeeger, J. Hummer, and F. Hanscom. *Operational Effects of Larger Trucks on Rural Roadways*. Presented at TRB Annual Meeting, January 1990).

Truck Exposure Data

A particular problem in truck accident analysis is the unavailability of truck exposure data. Truck exposure is not available by truck types and hence it becomes extremely difficult to study the impact of different types of trucks in an accident. In a recent report on data requirements for monitoring truck safety, it is emphasized that greater quality control is required in collecting truck data to get better truck exposure, especially by truck types (*5*).

A significant finding from previous research was the method of calculating truck accident rates. In a number of studies, truck accident rates are calculated by considering an accident a truck accident if at least one vehicle involved is a truck. The truck accident rate is determined by dividing the total number of truck accidents by truck annual daily traffic (ADT), resulting in artificially high truck accident rates. The reason behind this is that multivehicle accidents involving trucks and nontrucks are only counted as truck accidents. To get a true picture of truck accident occurrence, truck involvement rates should be used. That is, the total number of trucks involved in an accident divided by truck ADT.

Accident Models

A number of accident models have been developed for truck accidents exclusively. One of the truck models developed by Garber (*6*) established a loglinear relationship among truck involvement rate, slope change rate, ADT, and truck percentage variables. Three-year data were considered from Virginia and the road types used include divided and undivided, two- and four-lane, and primary and Interstate highways. In another attempt by the same author (*7*), an equation was developed to calculate the truck involvement using only truck volume. These are the only two significant models that deal with truck involvements.

In a study of truck accident modeling, Chang and Jovanis (*8,9*) have shown that truck accidents could be modeled at a disaggregate level as a survival process. Several truck accident models were developed using variables pertaining to truck characteristics, driver characteristics, roadway geometry and environmental factors. The significant variables in all of the models were weather condition, day and night condition, age and experience of the driver, the weight of the cargo carried, and the number of off and on duty hours worked by the driver.

In another study done by the Saccomanno and Buyco (*10*), a loglinear modeling approach is used to assess the effect of the traffic environment on truck accident rates. It should be noted that neither this model nor the previous work by Jovanis found any significant relationships between geometric variables and truck involvement rate.

Other models, although not developed exclusively for truck accidents, can provide a foundation for the development of a truck model. Zeeger et al. (*11*) developed a model that predicts single vehicle plus opposite direction head-on collisions, opposite direction sideswipe collisions, and same-direction sideswipe accidents/mi/year. The variables in this model included ADT, lane width, paved and unpaved shoulder widths, median roadside rating and terrain condition. The correlation coefficient (R^2) was 0.456. Data were collected from seven states on 8,350 km (5,000 mi) of two-lane rural roads.

In developing a relationship between rural highway geometry and accident rates in Louisiana, Dart (*12*) found that the percentage of trucks, traffic volume ratio, lane width, shoulder width, pavement cross slope, horizontal alignment, vertical alignment, percentage of continuous obstructions, marginal obstructions/mi, and traffic access points/mi were significant variables. The study was carried out on approximately 1,670 km (1,000 mi) of rural highway. Various models were developed and the model for total accidents had an R^2 of 0.46.

The studies noted previously are important as they have identified the variables that affect accidents and have narrowed down to a handful those that are most significant. The important variables that emerge from these studies are shoulder width, shoulder type, median width, median type, ADT, and lane width, supplemented by variables indicating the curvature and gradient of the roadway segment.

BRIEF DESCRIPTION OF HSIS

The data base used for the model development in this study is called HSIS. It has been developed by the FHWA and the University of North Carolina Highway Safety Research Center. This data base contains accident and roadway information collected over the period 1985 to 1989 from five states (Illinois, Maine, Michigan, Minnesota, and Utah). Although the data collected by each state are different, each state has three basic files: accident, roadway, and traffic. For location-based safety analysis, various files are combined using route-milepost as common reference variables.

Because different states in HSIS collect different variables, depending on the nature of the analysis, one or more states could be selected for use. For the present study, a preliminary analysis of the data base was done to determine the type and quality of the variables available in each state. The states were compared on the basis of the important variables identified in the literature review. The preliminary analysis indicated that Illinois and Utah have all the major variables required for the study, but Utah has more complete curvature and gradient variables. Hence, the state of Utah was selected for use in this study.

DATA ANALYSIS

Utah Accident File

The accident file for Utah contains approximately 37,000 accidents/year involving approximately 65,000 vehicles. There were 185,341 total reported accidents in Utah for the 5-year period (1985–1989), out of which 124,161 (70 percent) were property damage accidents, 44,178 (31 percent) were minor injury accidents, and 17,002 (9 percent) were serious accidents (combined fatal and incapacitating injury accidents). The overall accident rate was 1.72 accidents/km/year (2.87 accidents/mi/year) for the 5-year study period.

The important accident characteristics of the Utah data are shown in Table 1. The total number of trucks and nontrucks present in Utah accidents, along with their overall percentage with respect to total vehicles, is indicated in this table. The last two columns show the relative percentage of truck and nontruck involvements. These columns show that trucks are more involved in property damage accidents, serious accidents, daylight accidents, dry roadway condition accidents, run-off-road accidents, overturning accidents, sideswipe and single-vehicle accidents compared with nontrucks.

TABLE 1 Summary of Accident Statistics for Utah (Vehicle Based)

Variable Name	Total Number of		Total % of		Relative % of	
	Trucks	Non Trucks	Trucks	Non Trucks	Trucks	Non Trucks
Total Accs.	11060	327794	3.26	96.7	100	100
PDO Accs.	7762	218055	2.29	64.35	70.18	66.52
Injury Accs.	2074	81685	0.61	24.11	18.75	24.92
Serious Accs.	1224	28054	0.36	8.28	11.07	8.56
Daylight Accs.	8389	231998	2.48	68.47	75.85	70.78
Dark No luminaries.	1605	31555	0.47	9.31	14.51	9.63
Dark with luminaries.	5415	44244	0.16	13.06	4.89	13.50
Dawn or Dusk Accs.	454	16714	0.13	4.93	4.10	5.10
Dry Accs.	8336	244009	2.46	72.01	75.37	74.44
Wet Accs.	1189	46829	0.35	13.82	10.75	14.29
Snow Accs.	848	18816	0.25	5.55	7.67	5.74
Icy Accs.	626	16372	0.18	4.83	5.66	4.99
Muddy/Oily Accs.	23	476	0.01	0.14	0.21	0.15
Motor Vehicle	7519	269760	2.22	79.76	67.98	82.30
ROR	942	16952	0.28	5.00	8.52	5.17
Fixed & Other Obj	732	9223	0.22	2.73	6.62	2.81
ROR-Median	421	5815	0.12	1.72	3.81	1.77
Animals	386	11143	0.11	3.29	3.49	3.40
Overturn	362	1954	0.11	0.58	3.27	0.60
Ped/Bic	89	6411	0.03	1.90	0.80	1.96
Train	16	199	0.00	0.06	0.14	0.06
Single Vehicle	3063	49088	0.90	14.49	27.69	14.98
Rearend	2248	96658	0.66	28.53	20.33	29.49
Turning	1729	23253	0.51	6.86	15.63	7.09
Approach Angle	860	54994	0.25	16.23	7.78	16.78
Sideswipe-Pass	792	11532	0.23	3.40	7.16	3.52
Parked Vehicle	560	15055	0.17	4.44	5.06	4.59
Intersection	478	40560	0.14	11.97	4.32	12.37
Backing	384	6149	0.11	1.81	3.47	1.88
Passing	282	13740	0.08	4.06	2.55	4.19
Sideswipe-Opp	245	4908	0.07	1.45	2.22	1.50
Head-On	177	4787	0.05	1.41	1.60	1.46
Rearend-Pass	155	5705	0.05	1.68	1.40	1.74

Utah Roadlog and Traffic File

The roadlog file covers 21,710 km (13,000 mi) of roadway. Seventy percent of these are primary and the rest are secondary roads. Eighty-four percent of the mileage is for two-lane roads and 60 percent of the mileage has an annual average daily traffic (AADT) of less than 500. The traffic file has data on the curvature and gradient of roadway segments. The horizontal curvature file covers 9,719.4 km (5,820 mi) with variables indicating degree and direction of curvature, whereas the vertical grade file has 9,769.5 km (5,850 mi) of data with variables such as percent and direction of grade. The data used for the final model development were filtered out by eliminating the sections having a length of less than 1.67 km (1 mi). A section in the file was defined by the beginning and ending mile post and had various geometric variables attached to it. Sections with AADT less than 10; routes with zero truck percentage, indicating that there was no truck travel on them; and sections without curve or grade variables were eliminated to obtain the final file. The final file contained 2,073 sections covering 12,174 km (7,290 mi) of roadway.

The final file showed that most of the sections are in rural areas (97 percent) and very little roadway mileage is classified as local roads. Urban freeways and local roads have little for data 68.5 km (41.01 mi) out of 12,174 km (7,290 mi) because the majority of the roads in Utah are classified as either rural Interstates, or rural arterial collectors. Hence these categories were not considered for model development. Also 99 percent of approximately 10,688 km (6,400 mi) of roadway classified as primary arterials and arterial collectors is made up of two-lane roads. On the basis of these observations, two models were selected for development: Interstates and two-lane rural roads. The models developed for Interstates do not have shoulder width as an independent variable because the data for the Interstate highways indicated constant 3.05-m (10-ft) wide shoulders. The data used for the Interstate truck accident model included 264 road sections [1,200.24 km (718.71 mi)] and 1,787 total trucks involved in accidents. The two-lane rural road model included 1,614 road sections [10,458.8 km (6,259.17 mi)], with 1,313 total trucks involved in accidents.

MODEL DEVELOPMENT

Variables Used In Model

After the analysis of the Utah files and based on the conclusions of the literature review, the variables considered for model development were nontruck average annual daily traffic/lane (AADT), truck ADT/lane (TRUCKADT), shoulder width (SHLDWID), horizontal curvature and vertical gradient as the independent variables, and truck involvement rate/km/year (TINVOL/KM/Y) as dependent variables. Truck ADT is obtained by multiplying the truck percentage by AADT, whereas nontruck AADT per lane is determined by subtracting the truck ADT from the total AADT for each section. The truck percentage varied from 1 to 59 percent for roadway sections under consideration. In order to take the effect of the number of lanes in the model, both ADTs were divided by the number of lanes to obtain the AADT/lane. Other variables such as median width, median type, shoulder type, pavement width, and pavement type, which may be related to truck accident occurrence, were found to be incomplete in the Utah files and these were not considered in the model development.

The curvature and gradient file could not be directly linked to the roadway file because the beginning and ending mile post in the curve and grade file do not directly match with those of the roadway file. To take this fact into consideration, aggregate curve and grade variables were created, indicating the percentage of the road section having a particular percentage of grade or degree of curvature. Three categories for Interstates and four categories for rural two-lane roads were created on the basis of typical design guidelines (*13*). The variables for Interstates are HCUV1, HCUV2, and HCUV3 for degree of curvature between 1 and 2.5, 2.5 and 4, and ≥4, respectively, and GGRD1, GGRD2, and GGRD3 for the percentage of gradients between 1 and 3 percent, 3 and 5 percent and ≥5 percent.

One km of Interstate section is shown in Figure 1 to demonstrate the developed curvature and gradient variables. Assuming that the 1-km section has four 0.25-km subsections, each with curvature and gradients as shown, then the value for each variable will be HCUV1 = 25 percent, HCUV2 = 25 percent, HCUV3 = 0 percent, GGRD1 = 25 percent, GGRD2 = 0 percent and GGRD3 = 25 percent. A similar example for a two-lane rural road is shown in the lower half of Figure 1.

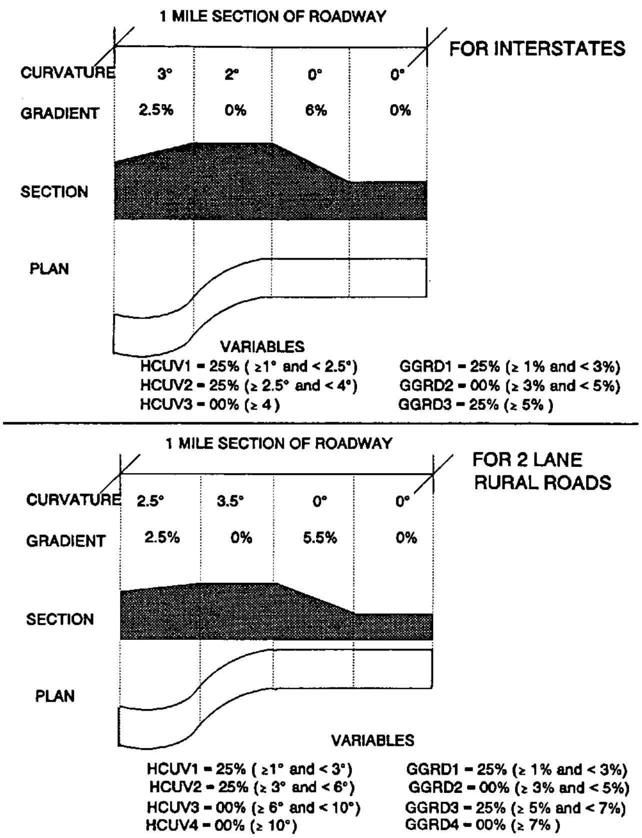


FIGURE 1 Curve and grade variables for Interstates and two-lane rural roads.

Selection of Models

Several general models identified in the literature were examined to determine their suitability for modeling truck involvement rates:

$$Y = \beta_0 (A_1)^{\beta_1} (A_2)^{\beta_2} (A_3)^{\beta_3} (A_4)^{\beta_4} \ldots \varepsilon \quad (1)$$

$$Y = \beta_0 + A_1\beta_1 + A_2\beta_2 + A_3\beta_3 + A_4\beta_4 \ldots + \varepsilon \quad (2)$$

$$Y = \beta_0 (\beta_1)^{A_1}(\beta_2)^{A_2}(\beta_3)^{A_3}(\beta_4)^{A_4} \ldots \varepsilon \quad (3)$$

where

β_0 = intercept,
$\beta_1, \beta_2, \beta_3, \beta_4$ = regression coefficients,
A_1, A_2, A_3, A_4 = geometric variables, and
Y = truck involvement rate/km/year.

A two-step process was used to determine the values of regression coefficients in these three models and to determine which model was best fitted using the available data. In the first step the stepwise SAS® [SAS is a registered trademark of SAS Institute, Inc., Cary, North Carolina.] procedure was used to determine which variables were significant at α = 0.05. The variables used in running the stepwise procedure for the Interstate model and the two-lane rural road model are shown in Table 2.

In the second step, the values of beta obtained from the stepwise procedure were used as initial values for the SAS® procedure NLIN, which is a nonlinear equation fitting procedure. This procedure produces the least-square estimates of the parameters through the Marquardt iterative method, where the residuals are regressed onto the partial derivatives of the model with respect to the parameters until the iteration converges. Using this procedure, it was found that there was no significant improvement in the parameter estimates obtained from the stepwise procedure, and the linear model had a higher R^2 compared with other models. Also, some of the coefficients for nonlinear models showed opposite signs than expected.

The best models, based on their R^2 values, are

Truck Accident Model for Interstates

$$\begin{aligned} \text{TINVOL}_i/\text{KM}/Y = & -0.1777 + 0.0002\text{AADT} \\ & + 0.0006\text{TRUCKADT} + 0.0053\text{HCUV2} \\ & + 0.0098\text{HCUV3} + 0.0022\text{GGRD2} \\ & + 0.0048\text{GGRD3} \end{aligned} \quad (4)$$

where $\text{TINVOL}_i/\text{KM}/Y$ is the truck involvement rate/km/year and R^2 equals 0.713.

TABLE 2 Variable Definition for Interstate and Two-Lane Rural Models

Variable Name	Definition
INTERSTATE MODEL	
$TINVOL_1$/KM/Y	Truck involvement rate per mile per year for Interstate
AADT	Average daily non-truck traffic per lane
TRUCKADT	Average daily truck traffic per lane
HCUV1	Percentage of roadway section with curve between 1° and 2.5°
HCUV2	Percentage of roadway section with curve between 2.5° and 4°
HCUV3	Percentage of roadway section with curve ≥ 4°
GGRD1	Percentage of roadway section with grade between 1% and 3%
GGRD2	Percentage of roadway section with grade between 3% and 5%
GGRD3	Percentage of roadway section with grade ≥ 5%
TWO LANE RURAL MODEL	
$TINVOL_2$/KM/Y	Truck involvement rate per mile per year for 2-lane rural road
AADT	Average daily non-truck traffic per lane
TRUCKADT	Average daily truck traffic per lane
SHLDWID	Shoulder width
HCUV1	Percentage of roadway section with curve between 1° and 3°
HCUV2	Percentage of roadway section with curve between 3° and 6°
HCUV3	Percentage of roadway section with curve between 6° and 10°
HCUV4	Percentage of roadway section with curve ≥ 10°
GGRD1	Percentage of roadway section with grade between 1% and 3%
GGRD2	Percentage of roadway section with grade between 3% and 5%
GGRD3	Percentage of roadway section with grade between 5% and 7%
GGRD4	Percentage of roadway section with grade ≥ 7%

Truck Accident Model for Two-Lane Rural Roads

$$\begin{aligned}\mathrm{TINVOL_2/KM/}Y = {} & 0.0027 + 0.00009\mathrm{AADT} \\ & + 0.0004\mathrm{TRUCKADT} - 0.0025\mathrm{SHLDWID} \\ & + 0.0011\mathrm{HCUV3} + 0.0007\mathrm{HCUV4} \end{aligned} \tag{5}$$

where $\mathrm{TINVOL_2/KM/}Y$ is the truck involvement rate/mi/year and R^2 equals 0.415.

All the variables in the models are significant at $\alpha = 0.05$. The variables AADT, TRUCKADT, CURVATURE and GRADIENT have positive signs, indicating that as the values of these variables increase, the truck involvement rate would increase. In Equation 4, the coefficients for HCUV3 and GGRD3 are larger than HCUV2 and GGRD2, respectively, showing that a road section with a degree of curvature ≥ 4 will lead to more accidents compared with degree of curvature between 2.5 and 4, and a section of a roadway with a gradient ≥ 5 percent will have higher truck accidents compared with a section with gradient between 3 and 5 percent. However, this is not true for the two-lane rural model according to Equation 5, and also it does not have any variables for gradient that may be caused by the inadequacy of the data. The shoulder width coefficient in the two-lane model has a negative sign indicating that with increased shoulder width, the truck involvement rate would decrease. The parameter estimates, along with the standard error values and t statistic for both the models, are shown in Table 3.

TABLE 3 Results of Interstate and Two-Lane Rural Highway Models

Variable	Parameter Estimate	Standard Error	t
MODEL FOR INTERSTATE HIGHWAY			
INTERCEPT	-0.1777	0.0577	-3.082
AADT	0.0002	0.0000	13.068
TRUCKADT	0.0006	0.0002	3.883
HCUV2	0.0053	0.0033	1.616
HCUV3	0.0098	0.0028	3.532
GGRD3	0.0022	0.0014	1.577
GGRD4	0.0048	0.0015	3.196
MODEL FOR TWO LANE RURAL HIGHWAY			
INTERCEPT	0.0027	0.0033	0.819
AADT	0.00008	0.0000	17.665
TRUCKADT	0.0004	0.0000	12.500
SHLDWID	-0.0025	0.0011	-2.286
HCUV2	0.0007	0.0003	2.324
HCUV3	0.0011	0.0005	2.485

For Interstate highway model:

$df = 6$; p-value = 0.000; Observations = 264, $R^2 = 0.731$

For Two lane rural road model:

$df = 5$; p-value = 0.000; Observations = 1,614, $R^2 = 0.415$

Model Validation

To validate the Interstate model, the data were divided in half using even- and odd-numbered observations. Even-numbered observations were used to develop the model for truck accident involvement for Interstates, and the following model resulted:

$$\begin{aligned} \text{TINVOL}_I/\text{KM}/Y = & -0.3014 \\ & + 0.0002\text{AADT} + 0.0009\text{TRUCKADT} \\ & + 0.0067\text{HCUV2} + 0.0139\text{HCUV3} \\ & + 0.0042\text{GGRD2} + 0.004\text{GGRD3} \end{aligned} \quad (6)$$

with $R^2 = 0.756$. The coefficients in this equation are close to Equation 4. This model was then used to determine the predicted value of truck involvement rates from the other half of the data. The predicted values and observed values matched closely with a correlation coefficient of 0.844 and the result of a t test on the values of predicted and observed values of truck involvement rate at $\alpha = 0.05$ also indicated that the validation was successful.

Employing a similar procedure for the two-lane rural roads model the following model resulted:

$$\begin{aligned} \text{TINVOL}_2/\text{KM}/Y = & -0.0025 \\ & + 0.00009\text{AADT} + 0.0004\text{TRUCKADT} \\ & - 0.0023\text{SHLDWID} + 0.0008\text{HCUV3} \\ & + 0.001\text{HCUV4} \end{aligned} \quad (7)$$

with $R^2 = 0.431$. The coefficient of correlation for predicted and observed truck involvement rate was 0.631 and in this case also t test at $\alpha = 0.05$ was successful in validating the model.

IMPLICATIONS OF THE MODELS

The two models developed could be used for comparing various sections of roadway and estimating the expected percentage decrease in accidents caused by geometric improvements. The following example illustrates the use of the models.

Consider a hypothetical 1-km section of Interstate roadway having an AADT of 4000, 5 percent trucks, 4 lanes, 3.05-m (10-ft) shoulders and 3.66-m (12-ft)-wide lanes. Let 40 percent of the section have 3 curvature, 60 percent have 6 curvature, 50 percent have 4 percent gradient, and the other 50 percent 6 percent gradient. Using Equation 5, the truck involvement rate can be calculated for this section. The truck involvement rates for the base section previously discussed and for the section obtained after certain modifications are shown in Table 4. Example 1 indicates that if the length of the 6 curve section is increased from 60 to 80 percent, the truck involvement rate increases. Similarly in Example 2, if the length of the 6 percent gradient section is increased from 50 to 70 percent, the truck involvement rate increases. Comparison of Examples 3 and 4 indicates that truck accidents are affected more by presence of curvature than gradient. Example 5 demonstrates the potential reduction in truck involvements if all curves and grades are eliminated.

CONCLUSION

In this paper the first step in modeling of the relationship between the roadway design variables and truck accident involvement rates is provided. Also pointed out is the difficulty in developing these models from the existing data sources. The lack of existing truck models in the literature can be directly traced to the lack of good data. This paper uses a newly developed data source in a location-based analysis to develop truck accident models. It provides an effective demonstration of both the potential of HSIS and the need to supplement this data with more accurate truck exposure and more detailed roadway design data.

The models developed illustrate the effect curvature and gradient have on truck accidents. For truck accidents on Interstates, the significant degree of curvature was found to be ≥2.5, whereas on two-lane rural roads it was ≥6. In the case of gradients for Interstates, the significant percentage was found to be ≥3 percent, whereas the two-lane rural model did not include a gradient variable. The appearance of the grade variable in the Interstate model with a high R^2 strongly suggests a lack of adequate data for the two-lane rural road model.

TABLE 4 Predicted Number of All Truck Involvement Rates Using Interstate Model

Examples	Non-Truck AADT/Lane	TRUCK AADT/Lane	HCUV2	HCUV3	GGRD2	GGRD3	TINVOL_I/M/Y
Base section	950	50	40	60	50	50	0.714
1	950	50	20	80	50	50	0.768
2	950	50	40	60	30	70	0.745
3	950	50	0	0	50	50	0.235
4	950	50	40	60	0	0	0.504
5	950	50	0	0	0	0	0.025

Conversion factor: 1 km = 0.6 mi.

ACKNOWLEDGMENTS

This study was sponsored by FHWA under the Graduate Research Fellowship Grants Program. That support notwithstanding, the opinions, findings, and conclusions presented are the responsibility of the authors.

REFERENCES

1. *Highway Fact Book, 1989.* Highway Users Federation and Automotive Safety Federation, Washington, D.C., 1989.
2. J. W. Hall and L. V. Dickinson. Truck Speeds and Accidents on Interstate Highways. In *Transportation Research Record 486*, TRB, National Research Council, Washington, D.C., 1974.
3. *Improving Truck Safety at Interchanges.* Technical Report, FHWA-IP-89-024, FHWA, U.S. Department of Transportation, Sept. 1989.
4. *Present Practices of Highway Transportation of Hazardous Materials.* Technical Report, FHWA-RD-89-013, FHWA, U.S. Department of Transportation, May 1990.
5. *Special Report 228: Data Requirements for Monitoring Truck Safety.* TRB, National Research Council, Washington D.C., 1990.
6. N. J. Garber and S. C. Joshua. *Estimating Truck Accident Rate and Involvements using Linear and Poisson Regression Model.* University of Virginia, Charlottesville, Published in the United Kingdom, 1989.
7. N. J. Garber and R. Gadiraju. Effects of Truck Strategies on Traffic Flow and Safety on Multilane Highways. In *Transportation Research Record 1256*, TRB, National Research Council, Washington D.C., 1990.
8. H-L. Chang and P. P. Jovanis. Formulating Accident Occurrence as a Survival Process. *Accident Analysis and Prevention*, Vol. 22, No. 5, 1990, pp. 407–419.
9. H-L. Chang and P. P. Jovanis. Disaggregate Model of Highway Accident Occurrence Using Survival Theory. *Accident Analysis and Prevention*, Vol. 21, No. 5, 1989, pp. 445–458.
10. F. F. Saccomanno and C. Buyco. Generalized Loglinear Models of Truck Accident Rates. In *Transportation Research Record 1172*, TRB, National Research Council, Washington, D.C., 1988, pp. 23–31.
11. C. V. Zeeger, R. C. Dean, and J. G. Mayes. Effect of Lane and Shoulder Widths on Accident Reduction on Rural, Two-lane Roads. In *Transportation Research Record 806*, TRB, National Research Council, Washington, D.C., 1981.
12. O. K. Dart and L. Mann, Jr. Relationship of Rural Highway Geometry to Accident Rates in Louisiana. In *Highway Research Record 312*, HRB, National Research Council, Washington, D.C., 1970, pp. 1–16.
13. *A Policy on Geometric Design of Highways and Streets.* American Association of State Highway and Transportation Officials, Washington, D.C., 1990.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Large-Truck Travel Estimates from the National Truck Trip Information Survey

DAWN L. MASSIE, KENNETH L. CAMPBELL, AND DANIEL F. BLOWER

The methodology of the National Truck Trip Information Survey conducted by the Center for National Truck Statistics of the University of Michigan Transportation Research Institute is described in this paper. The survey was conducted to achieve the two main goals of estimating the registered large truck population of the continental United States and providing detailed data on its annual mileage. Travel in the file can be cross-classified by road type, area type, and time of day, and broken down according to truck configuration, cargo body style, cargo type and weight, gross weight, number of axles, and driver characteristics. This type of detail is useful in risk assessment, because the risk of accident involvement depends on the operating environment as well as the physical characteristics of the truck.

As part of its continuing studies on the safety of large trucks, the Center for National Truck Statistics of the University of Michigan Transportation Research Institute (UMTRI) carried out a national survey of medium and heavy truck usage over a 15-month period from November 1985 to February 1987. Termed the National Truck Trip Information Survey (NTTIS), the work produced a wealth of unique data on the travel patterns of different types of large trucks. Described in this paper are the methodology and some of the findings of the NTTIS survey, expanding on the results presented in an earlier paper (*1*).

LARGE-TRUCK TRAVEL DATA

Reliable estimates of large truck travel are needed for many purposes. Government agencies require travel data for regulatory and policy decisions. Highway finance determinations and pavement damage assessments rely on mileage estimates. The trucking industry uses travel information to guide operations and safety management. Cost-benefit analyses of proposed safety countermeasures require accurate travel estimates. The focus of this paper is on the need for exposure information suitable for calculating accident rates of different truck configurations under various operating conditions. With such information, areas for improvement in truck safety can be identified, and, if addressed, the effectiveness of accident-reducing measures monitored (*2*).

Although the need for data on the annual travel of the U.S. large truck population is well established, meeting this need is a difficult task. Travel data collection differs greatly from accident data collection. Accidents are discrete events, whereas travel is a continuous process. All state police maintain records of accidents, enabling reliable estimates to be made of the incidence of large truck involvements. However, no comparable data are collected for truck mileage.

States do supply the Federal Highway Administration (FHWA) with travel estimates based on traffic counts, as part of the Highway Performance Monitoring System established by FHWA in 1978. This program involves federal, state, and local governments. The states estimate travel based on traffic counts taken along selected road sections (*3*). National mileage figures are produced for different road classes and types of vehicles. Various criticisms have been made of the FHWA mileage figures, however. The classification system for large trucks is coarse, distinguishing only combination vehicles from single-unit trucks. More problematic are criticisms of the estimating procedure itself. Mingo (*4*) describes a series of inaccuracies and inconsistencies at both the state and federal levels in producing FHWA mileage figures. Greene et al. (*5*) argue that FHWA estimates are based on a nonrandom sample of vehicle counts and that traffic counts themselves do not represent vehicle travel but merely traffic density at one point on a road.

A different approach to estimating truck travel is taken by the Truck Inventory and Use Survey (TIUS), conducted every 5 years by the Bureau of the Census. This survey is conducted via questionnaires mailed to a random sample of truck owners. Except for vehicle registration data on which the samples are based, all TIUS information is self-reported. The questionnaires concern the "typical" configuration and operation of trucks over a 1-year period. Consequently, TIUS produces overall travel estimates, but travel cannot be broken down according to operating environment or specific features of truck configuration. TIUS estimates are based on a robust sample of truck owners. The 1987 TIUS collected data on a total of 104,606 trucks, including light trucks.

NTTIS shares some similarities with TIUS, but there are also important differences. Like TIUS, most of the information in NTTIS was obtained through interviews with truck owners. In contrast to TIUS, however, travel information in NTTIS is based on actual trips made by truck drivers, not their characterizations of "typical" trips. Another strength of NTTIS is that it offers many details concerning truck configuration and operating environment. Travel can be cross-classified by road type, area type, and time of day, and trucks can be classified according to configuration, number of trailers, carrier type, cab style, fuel type, cargo body style, cargo type and weight, and weight, length, and number of axles of trailers and power units. The file also includes information on driver age and experience.

Center for National Truck Statistics, University of Michigan, Transportation Research Institute, 2901 Baxter Road, Ann Arbor, Mich. 48109-2150.

SURVEY DESIGN AND METHODOLOGY

The objective of the NTTIS was to estimate the number of large trucks in the United States and provide detailed data on their mileage. The survey was conducted through multiple telephone interviews with truck owners to collect data on the use of their vehicles on particular days. The resulting NTTIS file is a hierarchical data set consisting of three parts: a truck file, a truck-tractor trip file, and a straight truck trip file (*6*). The truck file contains vehicle, company (owner), and annual mileage information, with one record per vehicle. The tractor and straight truck trip files contain trip information, one record per trip, for each trip taken by a survey vehicle on a survey day. All three files include weight variables so that national truck population and travel estimates may be calculated.

Sampling Frame

The sampling frame for NTTIS was formed from the R.L. Polk files dated July 1, 1983. The Polk files describe all registered vehicles in the country, excluding pre-1973 model-year vehicles in California and all vehicles in Oklahoma. Hence, the NTTIS sampling frame reflected these omissions and excluded Alaska and Hawaii as well. The Polk files were extensively processed to eliminate duplicate registrations from state to state. Vehicles selected from the sampling frame were trucks with a gross vehicle weight rating (GVWR) greater than 10,000 lb. Excluded were all pick-up trucks (regardless of GVWR), all passenger vehicles (such as passenger vans, recreational vehicles, ambulances, and buses of any type), farm tractors, and government-owned trucks.

The sampling procedure treated each state as a separate stratum, and within each state, straight trucks were sampled separately from tractors. An UMTRI-developed algorithm was used to make power unit-type assignments for the sampling process. Sample sizes were specified for each state, roughly proportional to the size of its truck population, and an interval selection procedure was followed in each stratum. At least 30 straight trucks and 60 tractors were selected from each state, and California and Michigan were oversampled to increase the number of tractors that pull two trailers. A total of 8,144 trucks was selected from the Polk registration lists to form the sample for the survey.

Data Collection for the Truck File

Once the sample was drawn, the survey work was carried out in two phases (Figure 1). During the implementation phase, conducted from January to May of 1985, each truck selected in the sample was located and a description obtained. Survey interviewers tried to contact the most knowledgeable person available for implementation information. In the case of private persons, the best source was most often the owner. With large companies, contact people were typically fleet supervisors, dispatchers, mechanics, drivers, and so on. Once the initial contact was made, interviewers secured the owner's cooperation, confirmed the vehicle's identification, obtained descriptive information on the company and truck, including a recent odometer reading, and made arrangements for acquiring detailed mileage information on four random survey days. Survey interviewing was conducted by telephone whenever possible. Mail versions of the interview forms were used only when the interview could not be completed by phone.

Of the original sample of 8,144 vehicles, 564 or 6.9 percent were determined to be nonsample because they had either been destroyed, were no longer registered, had GVWRs under 10,000 lb, or were not trucks. Of the 7,580 remaining vehicles, interviews were completed for 6,305 cases, for a response rate of 83.2 percent. The other 1,275 cases were not completed, primarily because of problems in locating the owner. Refusals were encountered in only about 3 percent of the selected vehicles. Information on the 6,305 vehicles with completed interviews is contained in the NTTIS truck file, which includes 3,704 straight trucks and 2,601 tractors.

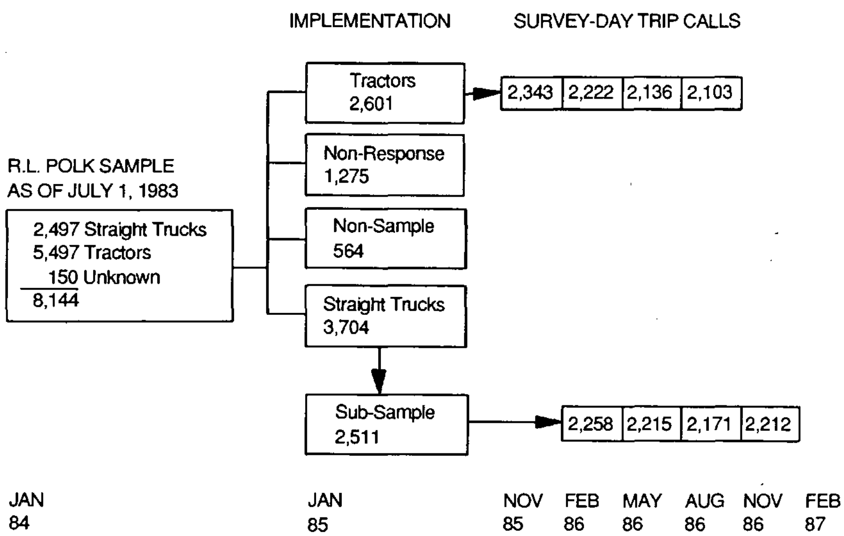


FIGURE 1 NTTIS case flow and timeline.

Data Collection for the Trip Files

After the implementation phase of the survey was complete, a sub-sample was drawn for the trip phase of 2,511 straight trucks and all 2,601 tractors. Most of the trip phase of the survey was devoted to collecting detailed information on the routes traveled by the selected vehicles and on the truck configuration, cargo, driver, and operating authority during these trips. Interviewers contacted truck owners quarterly over the course of a year to collect all trip information for a designated 24-hr period, usually the day before the phone call. Each "trip" lasted as long as the driver, operating authority, vehicle configuration, and cargo type and amount remained the same. Thus, if the driver changed, or cargo was loaded or unloaded, or one trailer type was exchanged for another, the interviewer began a new trip form to track the mileage put on by the new configuration. Tractor trip calls ran from November 3, 1985, through November 4, 1986, and straight truck calls were made from February 3, 1986, through February 5, 1987. A secondary goal of the trip phase was to collect a second odometer reading and usual or typical configuration data for all vehicles.

Vehicles selected for trip calls took a total of 13,097 trips, 4,966 by straight trucks and 8,131 by tractors. The trips were traced on specially prepared maps and the mileage broken down by road type, rural/urban, and day/night. The straight trucks traveled a combined sum of 206,276 mi, and the tractors logged 707,000 mi for an overall total of 913,276 mi. The value of the trip files lies in aggregating trip mileage across different travel categories for truck configurations of interest. The response rate for trip calls can be measured in two ways. Of the 5,112 vehicles selected for trip calls, some trip information was obtained, even if it was only that the vehicle was not in use, for 4,789 vehicles, for a response rate of 93.7 percent. It was hoped to complete four trip calls on each vehicle over the course of a year, for a total of 20,448 potential trip days. A total of 17,660 survey day cases was actually completed, for a survey day response rate of 86.4 percent. This rate was 88.2 percent for straight trucks and 84.6 percent for tractors. Overall, the in-use rate, that is, the percentage of vehicles that were actually used on the road on their survey date, was lower than anticipated. Straight trucks were in use on 27.0 percent of their survey days, and tractors were used at the slightly higher rate of 35.5 percent. The overall in-use rate was 31.3 percent, meaning that the typical vehicle was found to be in use on less than one-third of its survey days. In considering this apparently low usage, remember that NTTIS covered all registered trucks in the United States. This includes everything from trucks owned by Consolidated Freightways to farm trucks used mainly during the harvest season.

Mapping the Survey Trips

After a trip call was completed, research staff tracked the routes traveled on special maps prepared by UMTRI. The maps were based on the Rand McNally *Road Atlas* and followed its road type classification. Roads were divided into limited access highways, major arteries, and all other roads. Limited access roads include all U.S. Interstate highways, as well as state highways with fully controlled access. Major arteries include all U.S. and state routes that are not limited access, plus some other primary thoroughfares in large urban areas. All public roads that do not fall into the previous two categories make up the "other" road type group.

The special maps also included urban and rural zones. FHWA classifications were used to define three population categories: large urban areas (population of at least 50,000), small urban areas (population of 5,000 to 49,999), and rural areas. Local and county-wide maps showing the FHWA urban areas were obtained from each state so that exact boundaries could be marked on the *Road Atlas* maps. This made it possible to map the portion of the trip mileage in each of the three population areas precisely.

In addition to road type and population area, trips were broken down according to daytime and nighttime mileage. Because it was not feasible to ascertain the actual point on a trip where dawn or dusk came, "daytime" was arbitrarily set as 6 a.m. to 9 p.m. and "nighttime" as the 9 remaining hours. Therefore, nearly all of the travel classified as "night" was driven during darkness, but a small portion of the travel classified as "day" was actually driven in the dark, depending on the season of the year.

Adjustment Factors

A number of adjustment factors were calculated to correct for missing data encountered at several of the stages of data collection [for a full description, see Blower and Pettis (*6*)]. One important adjustment factor concerns mileage. As will be discussed in more detail later in this paper, total annual travel was estimated both from the information collected on the survey days and from two odometer readings obtained during the survey year. Estimates from odometer readings indicated greater annual travel than survey day estimates. Because odometer readings appear more accurate, an odometer adjustment factor inflates the mileages obtained from aggregating survey day travel to the mileages shown by odometer readings.

File Applications

NTTIS was designed to be a reliable sample of the real-world operating experience of trucks on the road. The data were collected on the basis of actual trips made by large trucks and can be used to produce national population and mileage estimates. A major application of the NTTIS travel file is to estimate the risk of large truck accident involvement under particular conditions. Large trucks are themselves a heterogeneous group, varying widely in size, configuration, and cargo, and they travel under many different circumstances. They operate on different classes of roads, in areas of varying population density, traveling at all hours of the day and night. All of these factors may influence the risk of accident involvement.

The NTTIS file allows truck travel to be cross-classified by many factors of interest. Every survey trip can be characterized in terms of day and night miles over three road types and three population types. By aggregating different types of travel across trips and survey days, annual mileage estimates can be produced for particular truck configurations. For example, mileage distributions can be compared between tractors hauling a van semitrailer and tractors with a flatbed trailer. The total annual mileage of these two configurations can be calculated, as can the proportion traveled on different road types or during the daytime versus the nighttime. By combining this information with the number of annual accident involvements for these configurations, the actual risk of accident involvement under the particular conditions may be estimated.

AVERAGE ANNUAL MILEAGE ESTIMATES IN NTTIS

The NTTIS file contains three independent estimates of average annual mileage. The first is the owner's estimate of annual travel, which is referred to here as "self-reported" annual mileage. The second is calculated from odometer readings supplied for specific dates near the beginning and end of the 1-year trip survey period. The third estimate is derived from travel reported on the individual survey days inflated by the selection weights for these dates. A comparison of the three estimates by power unit type (Figure 2) shows that the self-reported figures are the highest and the mileage from the survey days the lowest. An evaluation of these differences requires an understanding of the procedures used to obtain each measure of travel.

Deriving Average Annual Travel

When the truck owners were first contacted during the implementation phase of the survey, interviewers asked them to estimate how far they planned to drive the power unit over the following 12 months. An estimate based on the previous 12 months was acceptable if they planned to use the power unit in the same way. The self-reported figures are the highest of the three NTTIS travel estimates, averaging 55,149 mi for tractors and 12,547 mi for straight trucks. It is possible that owners sometimes overestimate the annual travel of their trucks. Because the estimate is made on the spot in the course of a telephone interview, the owner may not consider factors that lower a power unit's actual annual mileage from its planned use. Such factors include basing the estimate on high-mileage days instead of "average" days, not considering the time a power unit is out of service for maintenance and repairs, and not taking into account the rotating use of tractors in trucking operations.

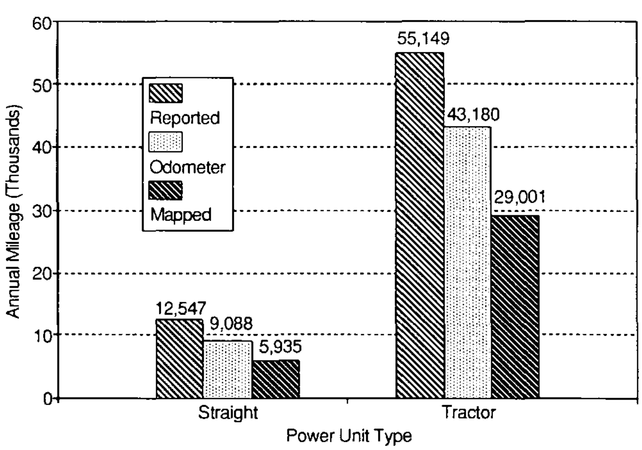


FIGURE 2 Average annual mileage in NTTIS by data source.

The second means of deriving annual travel was to annualize the two odometer readings. The odometer-based derivations average 43,180 mi for tractors and 9,088 mi for straight trucks. Although these figures are about 20 to 25 percent lower than the self-reported mileage estimates, they might be expected to be more accurate simply because they are a more objective measure. The main problem with the use of the odometer figures in the NTTIS file is that two readings were not obtained for more than 40 percent of the trucks included in the trip survey. This reflects the difficulty involved in obtaining odometer readings. Accurate figures require contacting the respondent at two specific times during the year, and problems result if the power unit is not present when the calls are made or if the odometer has been broken or changed during the course of the year.

The third procedure for calculating average mileage was based on the travel information collected on the four survey days. Researchers tracked the actual routes followed by a vehicle for each 24-hr survey period and totaled and annualized the mileages. The mapped annual mileages turned out to be about one-third lower than the odometer readings, averaging 29,001 mi for tractors and 5,935 mi for straight trucks. Because the proportion of trucks reported not to be in use on the survey days was rather high, it is likely that sometimes trucks were reported incorrectly as not in use.

Discussion of Differences between Mileage Estimates

Part of the difference among the three types of mileage estimates in NTTIS is related to the timing in obtaining the estimates. Self-reported mileage estimates essentially pertain to the year 1985, whereas odometer and mapped miles roughly describe travel during 1986. Because truck mileage generally declines with the age of the truck, the self-reported estimates would be expected to be somewhat higher than the odometer and mapped estimates, because the former describe a population that is about a year younger than the latter.

It is possible to estimate the effect of the time lag between the self-reported and odometer miles. Average annual mileage was plotted by model year for self-reported miles, odometer miles, and odometer miles shifted by one year, to bring those estimates in line with self-reported estimates for time. Separate plots were prepared for straight trucks and tractors. Regression lines were plotted through each set of points, and the average distance between the lines for self-reported miles and unshifted odometer miles was calculated. The average distance between the lines for self-reported miles and shifted odometer miles was also calculated. The results indicated that the year's difference in timing explains 19.4 percent of the difference between self-reported and odometer estimates for straight trucks and 26.7 percent of the difference for tractors.

The self-reported average annual mileage figure for all straight trucks in NTTIS is 12,547 mi, and the odometer estimate is 9,088 mi, a difference of 3,459 mi. Assuming that 19.4 percent of this difference is caused by the time lag, the new odometer estimate would be raised to 9,760 mi, 2,787 mi below the self-reported figure. For NTTIS tractors, the average self-reported estimate is 55,149 mi and the average odometer estimate is 43,180 mi, a difference of 11,969 mi. Attributing 26.7 percent of this difference to the time delay results in a new odometer estimate of 46,375 mi, which is 8,774 mi under the self-reported figure.

Thus, the difference in time coverage probably accounts for about 20 to 25 percent of the difference between the self-reported and odometer estimates in NTTIS. Obviously other factors are also involved in the differences among the three types of estimates. The fact that three methods of calculating average annual mileage have yielded three different mileage estimates underscores the point that estimating truck travel is a very difficult task.

There is good reason to think that annual travel estimates by truck owners are too high. The owner is asked to provide an estimate for the entire year. It is unlikely that down time for repairs, accidents, or normal rotation of vehicles within a fleet will be considered. Travel estimates based on trip calls are almost certainly too low. Some travel was undoubtedly not reported, either inadvertently or to limit time spent on the interview. In any case, measurement error from trip calls is biased toward underreporting, because it is more likely that trips were overlooked than invented. Odometer readings provide a more objective, reliable means of estimating annual travel, although it is conceded that they too are subject to error. Despite extensive efforts, two odometer readings were obtained for just 58.6 percent of trucks selected for trip calls, a missing data rate significantly higher than for any other data element in NTTIS. It is possible that nonresponse bias affects the accuracy of the odometer estimates, although the direction of this effect is unknown. Trucks with high annual mileages may be more likely to be unavailable for odometer readings because they are on the road. Alternatively, little-used trucks may be inaccessible for different reasons, thereby raising the overall odometer estimates. Even with this uncertainty, odometer readings provide the best estimate of overall travel. Accordingly, mapped miles from survey calls are weighted by the odometer estimates.

TRAVEL DISTRIBUTIONS

Travel patterns of trucks, in terms of total travel and the distribution of road type, area type, and day/night, vary with respect to many of trucks' physical features. Power unit type, configuration, gross combination weight (GCW), and number of axles all are associated with different travel patterns. In this section, mileage distributions across the three travel categories will be examined for specific truck types of interest in order to illustrate some of the differences that exist. [For additional travel distributions based on NTTIS data, see Massie et al. (*7*).] The distributions in this section are based on mileage estimates from the mapped trips, inflated by the odometer adjustment factor.

Truck Configuration and Operating Environment

Large truck travel varies a great deal according to power unit type and configuration. Based on NTTIS estimates, straight trucks outnumber tractors in the national large-truck population by about 70 percent to 30 percent. NTTIS estimates a national population of 2,185,630 ± 26,063 straight trucks and 919,702 ± 26,736 tractors. The distribution nearly reverses for annual travel, however, with tractors logging 68 percent of the total miles and straight trucks only 32 percent (37,870 million mi ± 695 million for tractors and 17,990 million ± 513 million for straights). This is because the average annual mileage of a tractor is about five times that of a straight truck (41,176 to 8,231 mi). Trucks can also be broken down in NTTIS according to configuration, such as straight trucks alone, straight trucks hauling one or two trailers, bobtails (tractors alone), singles (tractors hauling one trailer), and doubles (tractors hauling two trailers). In this section, tractors with three trailers, or triples, will be included with doubles because they are made up of such a small category. NTTIS estimates that a total of about 55,560 million mi is logged annually by the five main large truck configurations. Singles accumulate 35,010 million mi ± 689 million each year (63 percent of the total), and straight trucks with no trailers are next with 16,680 million ± 485 million mi (30 percent). Bobtails, doubles, and straight trucks pulling trailers together account for only 7 percent of the total large-truck travel.

Breaking down truck travel by road class, there are marked differences among configuration types. The distribution of each configuration's miles over the three road classes is illustrated in Figure 3. The proportion of limited access travel ranges from less than 20 percent for straight trucks to more than 72 percent for doubles. Conversely, travel on major arteries drops from 42 percent for straight trucks to 20 percent for doubles, and mileage on all other types of roads ranges from 38 percent for straight trucks to less than 8 percent for singles and doubles. These distributions provide an example of the importance of considering factors in addition to total travel when calculating the risk of accident involvement. Limited access routes are generally much safer than other types of roads (*8*). Therefore, vehicles like singles and doubles that

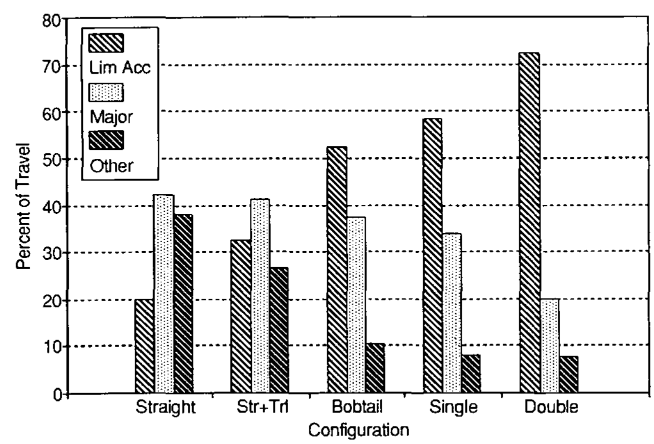


FIGURE 3 Travel by configuration and road type.

log a large proportion of their travel on the Interstates are exposed to less accident risk per mile than straight trucks, which travel much less frequently on limited access routes.

Considering travel on rural versus urban roads (following FHWA classifications), there are again substantial differences in the distributions among configurations (Figure 4). Single-unit straight trucks log approximately equal numbers of miles in rural and urban areas, whereas tractor-semitrailers put on more than twice as many rural as urban miles. Turning to the third main travel factor, time of day, this breakdown by configuration type can be seen in Figure 5. All five configurations put on far more miles during the day than at night, but again the proportions vary. Straight trucks accumulate less than 3 percent of their miles at night, whereas the nighttime portion is nearly 19 percent for singles and more than 34 percent for doubles. Just as for road class, area type and time of day are travel factors that affect a vehicle's risk of accident. As can be seen in Figures 3 to 5, mileage distributions over all three of these factors vary significantly from one truck configuration to another.

NTTIS also can be used to generate mileage estimates for combinations of the factors treated previously. Eight categories can be defined by generating all combinations of two road types (limited access versus all other roads), two light conditions (day versus night), and two area types (rural versus urban). If the aggregate travel distribution for the five truck configurations across the eight categories of travel is prepared, the category with the largest share of the mileage, at about 30 percent, is rural "other" roads during the daytime. The next largest is rural limited access roads during the daytime, with 22 percent. Urban limited access roads during the daytime represent 15 percent of travel, rural limited access roads at night 7 percent, and urban limited access roads at night 3 percent. Urban "other" roads during the day account for 18 percent of travel, rural "other" roads at night 3 percent, and urban "other" roads at night 1 percent. There is clearly more travel during the day than at night, particularly on "other" classes of roads. There is also more travel in rural compared with urban areas, and on "other" roads compared with limited access roads, although this last difference is not as great.

The travel distribution over the eight categories is shown separately for straight trucks with no trailers, singles, and doubles in Figure 6. Straight trucks accumulate much more travel on "other" roads, compared with singles and doubles, and put on very little nighttime mileage. Singles, on the other hand, accumulate substantial travel on limited access roads and have a higher proportion of night travel than straight trucks. Most of the doubles travel is on limited access roads, in part because of restrictions in some states, but a large share of their travel is also at night. Doubles are operated more uniformly around the clock and are used primarily in long-haul, general freight operations.

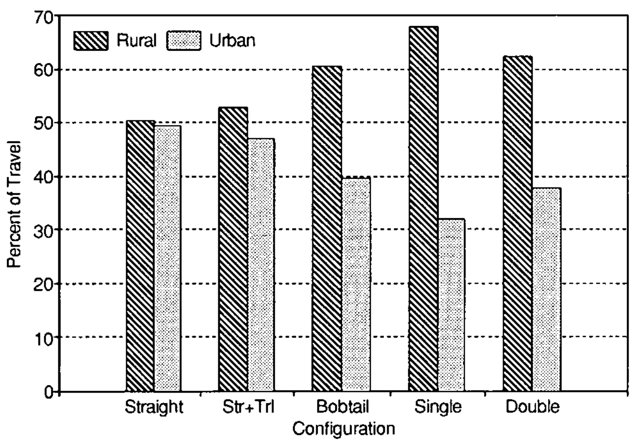


FIGURE 4 Travel by configuration and area type.

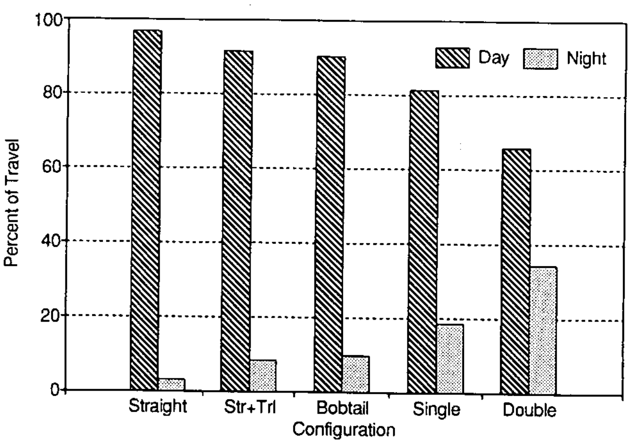


FIGURE 5 Travel by configuration and time of day.

Gross Combination Weight (GCW)

The comparisons discussed so far have classified large trucks according to power unit type and configuration. NTTIS can also show travel by the actual GCW of the vehicle. Presented in Figure 7 are the travel distributions of straight trucks and tractors in 10,000-lb increments of GCW. The category labels for this and subsequent figures are for the lower bound of the GCW increment. So, for example, the bars labeled "20" rep-

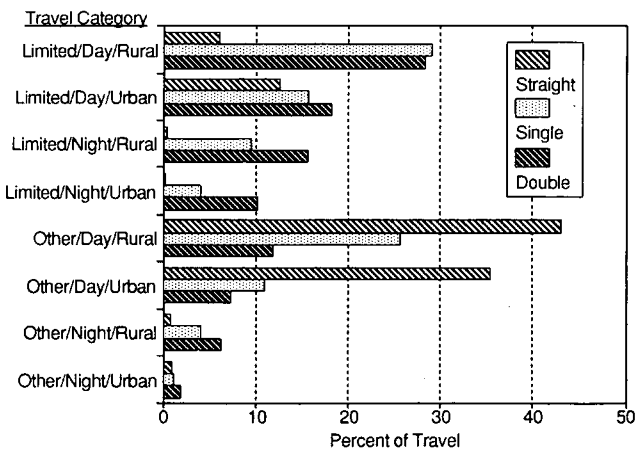


FIGURE 6 Travel by road type, time of day, and area type for straight trucks, singles, and doubles.

resent GCWs of 20,000 to 29,999 lb. Cases with missing data on GCW have been excluded from the distributions.

The operating characteristics of the two power unit types are quite different and the differences are reflected in their operating weights (Figure 7). Most straight trucks are Class 6 or below and operate without trailers. Accordingly, more than half of the travel of straight trucks is at weights under 20,000 lb, and more than three-quarters is at weights under 30,000 lb. The travel at higher weights reflects in large part the operations of loaded Class 7 and 8 straight trucks and straights with trailers. Tractors, in contrast, are primarily Class 7 and 8 and operate over 95 percent of the time with trailers, typically one. The empty weights of singles and doubles are roughly comparable, and the peaks of the bimodal distribution of tractor travel in Figure 7 reflect empty and loaded combinations.

Considering GCW for loaded vehicles only, the travel distributions naturally change, but the distinction between straight trucks and tractors remains clear. As a group, straight trucks travel slightly more without any cargo than do tractors. NTTIS estimates that about 36 percent of straight truck miles are in an unloaded condition, compared with only 30 percent for tractors (including bobtails). Illustrated in Figure 8 is travel according to GCW for large trucks that are at least partially loaded. This figure may be compared with Figure 7 to see the effect of excluding empty vehicles. The change in the GCW travel distribution for straight trucks is relatively minor. The under-20,000-lb class has dropped from 51 percent to 43 percent of the overall mileage, and all of the heavier categories show a slight rise as a result. This is evidence of the dominance among straights of weight classes up to Class 6, which are rated at no more than 26,000 lb. In contrast, the tractor distribution has changed substantially with the exclusion of the empty vehicles. The peak at the 20,999 to 29,999-lb category has disappeared, whereas the peak at the 70,999 to 79,999-lb class has risen. More than 58 percent of loaded tractor travel occurs at a GCW of 60,000 lb or greater, and 43 percent is conducted at a GCW of at least 70,000 lb. The comparable figures for loaded straight trucks are 7.5 percent and 4 percent, respectively.

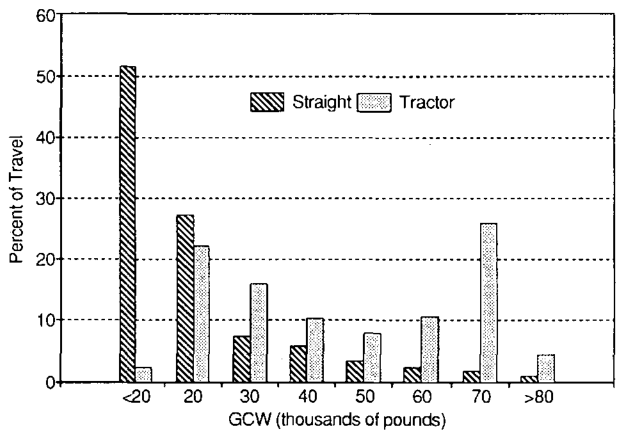


FIGURE 7 Travel by power unit type and gross combination weight.

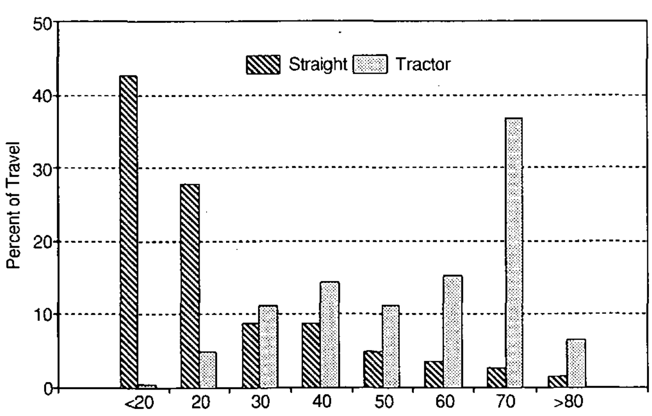


FIGURE 8 Travel by power unit type and gross combination weight, loaded trucks only.

Axle Configuration

Axle configuration is another large-truck characteristic that was considered in the NTTIS survey. The number of axles on each unit of a configuration was recorded, and if this number changed, as when a lift axle was raised or lowered, a new trip form was started. Thus, NTTIS contains the same detailed mileage information according to axle configuration as that already described for configuration type and GCW.

Provided in Figure 9 is an overview of tractor-trailer travel according to number of axles. The first six bars on the graph pertain to singles and the last two to doubles and triples. In each case, the first number indicates the number of axles on the tractor and the next one (or two) the number of axles on the trailer. Represented by "O/O" and "O/O/O" are "other" axle combinations for singles and doubles, respectively. By far the most common configuration for singles is a 3-axle tractor hauling a 2-axle trailer. This axle configuration accounts for nearly 74 percent of all tractor-trailer travel. The next most common configuration for singles is the 2/2 combination, which accounts for 11 percent of all tractor-trailer mileage. Among doubles, 2/1/2 is the most common configuration. This combination represents about 60 percent of all multi-trailer travel and 3 percent of overall tractor-trailer mileage.

In Figure 10, travel distributions are compared according to GCW for 3/2 and 2/1/2 axle configurations, which are the most common configurations for singles and doubles, respectively. The 3/2 singles have a greater share of travel at both ends of the GCW scale than do the 2/1/2 doubles. More than 37 percent of 3/2 singles travel occurs at a GCW of at least 70,000 lb. This compares with less than 27 percent of the 2/1/2 doubles travel. However, the doubles drive more of their miles in the very heaviest GCW class 80,000 lb and over than do the singles, 6.5 percent to 4 percent. Travel at GCWs under 40,000 lb accounts for more than one-third of the 3/2 singles mileage but only a quarter of the 2/1/2 doubles mileage. The higher proportion of travel at low GCWs for singles is likely caused by a typically lower empty weight compared with doubles. The greater share of travel at high GCWs may be caused by 3/2 singles frequently hauling higher-density cargo than 2/1/2 doubles, which are used for general freight.

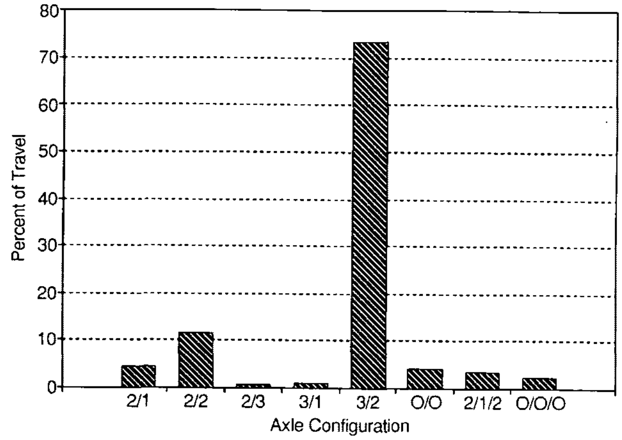


FIGURE 9 Tractor travel by axle configuration.

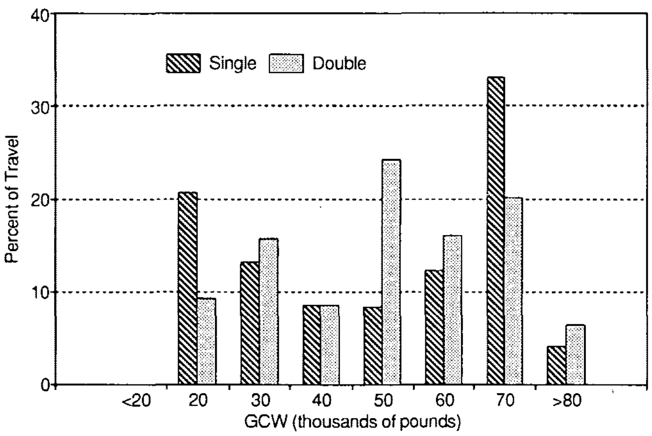


FIGURE 10 Travel by gross combination weight for 3/2 singles and 2/1/2 doubles.

CONCLUSIONS

The series of comparisons presented in the last section illustrate several important aspects of the national large-truck travel experience. The first is that different types of trucks have substantially different distributions of travel across categories defined by road class, time of day, and population area. Because these categories of travel are associated with different risks of accident involvement, the travel patterns of any given type of truck have a strong influence on the likelihood that a truck of that type will be involved in an accident. Second, large trucks form an extremely heterogeneous group. This is reflected in travel comparisons that consider power unit type, GCW, and axle configuration. Large trucks vary widely in their physical configuration, and this also has a bearing on the risk of accident involvement.

The diversity of trucking operations underscores the importance of reliable travel data in any analysis that seeks to determine the relative safety of one truck type versus another. To carry out the analysis, it is essential to have both accident data and travel data that can be cross-classified by the factors of interest, especially those categorizing the type of travel. It is not sufficient simply to know the total miles traveled. It must also be possible to classify the travel by factors related to accident risk, such as type of road and time of day.

NTTIS meets these criteria for reliable, detailed large-truck travel estimates, but the current file is already becoming outdated. Whereas UMTRI's Center for National Truck Statistics has been conducting a survey of large trucks involved in fatal accidents since 1980, NTTIS was a one-time project that surveyed truck use in 1985 to 1986. The U.S. trucking industry is a dynamic one that changes along with the economy, demographics, size and weight legislation, truck equipment and configurations, technology, traffic densities, as well as the nature of the highways on which trucks operate with other vehicles. Reliable truck travel information is a continuing need. Truck safety continues to be a matter of major national importance. To meet the demonstrated need of reliable, current estimates of heavy truck travel, NTTIS should be conducted on a regular basis, ideally once every 2 years.

ACKNOWLEDGMENT

This analysis of the National Truck Trip Information Survey was supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.

REFERENCES

1. K. L. Campbell. Population Estimates from the National Truck Trip Information Survey. In *Transportation Research Record 1068*, TRB, National Research Council, Washington, D.C., 1986, pp. 76–84.
2. *Special Report 228: Data Requirements for Monitoring Truck Safety.* TRB, National Research Council, Washington, D.C., 1990.
3. *Highway Statistics 1990.* FHWA, U.S. Department of Transportation, 1991.
4. R. D. Mingo. *Evaluation of FHWA's Vehicle Miles of Travel Estimates for Heavy Vehicles.* Intermodal Policy Division, Association of American Railroads, 1991.
5. D. L. Greene, P. S. Hu, and G. F. Roberts. Analysis of Geographical and Temporal Variation in Vehicle Classification Count Statistics. In *Transportation Research Record 987*, TRB, National Research Council, Washington, D.C., 1984, pp. 21–28.
6. D. F. Blower and L. C. Pettis. *National Truck Trip Information Survey.* Report UMTRI-88-11. University of Michigan Transportation Research Institute, Ann Arbor, 1988.
7. D. L. Massie, K. L. Campbell, D. F. Blower, and A. C. Wolfe. *Large Truck Travel Estimates from the National Truck Trip Information Survey.* Report UMTRI-91-39. University of Michigan Transportation Research Institute, Ann Arbor, 1991.
8. K. L. Campbell, D. F. Blower, R. G. Gattis, and A. C. Wolfe. *Analysis of Accident Rates of Heavy-Duty Vehicles.* Report UMTRI-88-17. University of Michigan Transportation Research Institute, Ann Arbor, 1988.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Comparison of Large-Truck Travel Estimates from Three Data Sources

DAWN L. MASSIE, KENNETH L. CAMPBELL, AND DANIEL F. BLOWER

The number of miles traveled each year by the U.S. large-truck population is a topic of interest for many reasons, one of which is safety. Although the number of accidents involving large trucks may be easily calculated from accident data, it is often more informative to know their risk of accident involvement per mile of travel. This requires accurate travel data. Compared in this paper are three sources of truck travel data: the Truck Inventory and Use Survey conducted by the Bureau of the Census; the National Truck Trip Information Survey conducted by the University of Michigan Transportation Research Institute; and annual estimates published in *Highway Statistics* by the Federal Highway Administration. Each data source yields different estimates of annual travel by large trucks, which is to be expected considering the difficulty of collecting travel data. The overall conclusion, however, is that the Truck Inventory and Use Survey and the National Truck Trip Information Survey estimates are reasonably close to each other, whereas *Highway Statistics* estimates are significantly higher. The implication of this finding is that the procedures used by the states and the Federal Highway Administration to generate *Highway Statistics* data lead to artificially and systematically high estimates of travel by large trucks.

The Center for National Truck Statistics of the University of Michigan Transportation Research Institute (UMTRI) conducted a national survey of medium and heavy trucks beginning in January of 1985. Termed the National Truck Trip Information Survey (NTTIS), the study produced estimates of the national registered large-truck population and its annual travel. The methodology of NTTIS is described in detail in a companion paper (*1*). In this paper, estimates of large-truck travel from NTTIS and two other sources are compared. One purpose of the comparisons is to assess the degree of correspondence among the three travel estimates. Another is to illustrate the inherent difficulty in measuring truck travel and the benefit of considering multiple sources of travel data. Despite the difficulty and associated cost of collecting travel data, truck travel information is vitally needed in order to make informed decisions on a host of topics, particularly those concerning truck safety.

NTTIS AND TIUS COMPARISONS

The comparisons start with two sources of data, NTTIS and the Truck Inventory and Use Survey (TIUS). TIUS is conducted every 5 years by the Bureau of the Census as part of the Census of Transportation. NTTIS and TIUS begin with a common base, the R.L. Polk vehicle registration files. The sampling frame for NTTIS was formed from the July 1, 1983, Polk files. The two most recent TIUS surveys were drawn from the July 1, 1982, and July 1, 1987, versions of the Polk files respectively. NTTIS restricted its sample to trucks with a gross vehicle weight rating (GVWR) greater than 10,000 lb. All pick-up trucks, regardless of GVWR, were excluded from the sample, as were all passenger vehicles (such as passenger vans, recreational vehicles, ambulances, and buses of any type), farm tractors, and government-owned trucks. Similar to NTTIS, TIUS excluded ambulances, open utility vehicles, motor homes, buses, farm tractors, and government-owned vehicles. Unlike NTTIS, TIUS sampled trucks of any GVWR, including light trucks.

The implementation phase of NTTIS was carried out in January through May of 1985. As part of this phase, survey interviewers contacted truck owners and asked them how far they drove their power unit in a year. Phone interviewers also obtained descriptive information on the truck and the company at this time. The implementation phase produced data on 6,305 trucks. During the subsequent trip phase of NTTIS, truck owners were contacted by phone four times over the course of a year. Each time, interviewers sought information on all trips made by the truck in a specific 24-hr period. Detailed physical information on the truck and its cargo was collected, and the routes traveled by the truck were mapped according to road type, population area (rural/urban), and time of day (day/night).

TIUS is conducted through survey forms mailed to owners of selected trucks beginning in January of the year after the Polk sample is drawn. Owners characterize their trucks in terms of the typical configuration and use over the previous year. This includes an estimate of the number of miles traveled, as well as information on the number of trailers usually hauled, type of cargo usually carried, typical weight of a load, and so on. The 1982 TIUS collected data on a total of 84,334 trucks, including light trucks. The 1987 TIUS total was 104,606 trucks.

Before the completion of NTTIS, TIUS had been the only national data base concerning the use of large trucks. Therefore, it is important to consider whether major differences exist between NTTIS and TIUS, given that NTTIS sampled a smaller proportion of the national truck population than TIUS. Whenever possible, NTTIS data elements were designed to be compatible with TIUS in order to facilitate comparison between the two. This section will compare truck population and travel estimates derived from NTTIS with those from the two TIUS surveys.

Center for National Truck Statistics, University of Michigan, Transportation Research Institute, 2901 Baxter Road, Ann Arbor, Mich. 48109-2150.

Truck Population

Estimates of the registered large-truck population in the continental United States by power unit type can be derived from both NTTIS and TIUS. The number of straight trucks is estimated at 2,534,973 by 1982 TIUS; 2,185,630 by NTTIS; and 3,230,210 by 1987 TIUS. The NTTIS estimate is about 14 percent lower than 1982 TIUS and 32 percent lower than 1987 TIUS. The number of truck-tractors is 900,884 according to 1982 TIUS; 919,702 according to NTTIS; and 1,038,130 according to 1987 TIUS. NTTIS estimates about 2 percent more tractors than 1982 TIUS and about 11 percent fewer tractors than 1987 TIUS.

At least three factors affect the degree of correspondence among the estimates from the three files. One is that the samples were drawn from three different registration years. Generally one would expect a small increase in the number of registered trucks from year to year, assuming favorable economic conditions. The other two factors are more complex and will be discussed in the next few paragraphs. One concerns identifying medium- and heavy-duty trucks in the TIUS data, and the other involves the time gap in NTTIS between drawing the sample and conducting the survey.

Large Trucks and GVWR

From the outset, the NTTIS survey was restricted to medium- and heavy-duty trucks, those with a GVWR over 10,000 lb. In contrast, TIUS samples all trucks, including light trucks. GVWR is encoded in the Vehicle Identification Number (VIN) for almost all trucks manufactured after 1980. R. L. Polk has developed decoding algorithms to extract this information from the VIN, and this code was included in the data supplied for the NTTIS survey. The Polk-derived GVWR is also included in the 1982 TIUS file but not in the 1987 version.

The VINs of some trucks, particularly those from model years before 1981, do not directly contain the GVWR. For many of these cases, the Polk-derived GVWR is based on the truck model as derived from the VIN, with the highest GVWR available for that model (as an option, for example) assigned. For many specific models, the majority of sales are at lower GVWRs. To improve the accuracy of the 10,000-lb GVWR cutoff when the NTTIS sample was drawn, UMTRI specified whether particular models should be included or excluded, in some cases overriding the Polk-derived GVWR. Models and series were identified for inclusion or exclusion based on sales information provided by the manufacturers. If the manufacturers indicated that the majority of sales were at a GVWR of 10,000 lb or less, then all of that specific model and series were excluded. The objective was to prevent the inclusion of an entire series when only a small fraction was actually rated over 10,000 lb. The models most influenced by this procedure were small step vans and pick-up truck models sold as a cab and chassis. The latter often have a flatbed or stake body added. To further ensure accuracy, GVWR was confirmed with the owner during the implementation phase of NTTIS.

Restricting the sample to trucks with GVWRs of more than 10,000 lb was not an issue for TIUS because light trucks are included in that survey. The Polk GVWR can be used to identify large trucks in the 1982 TIUS file, but for the reasons just stated some light trucks probably receive a Polk-derived GVWR over 10,000 lb. This would increase population estimates of medium-duty trucks, primarily straight trucks. The situation is worse for the 1987 TIUS file because that version does not include a GVWR variable. The file contains an average gross vehicle weight (GVW) variable based on the owner's estimate of the average weight of the vehicle when carrying a typical payload. GVW is only loosely related to GVWR, however, and rejecting all cases with average GVW below 10,001 lb would result in the exclusion of many medium-duty trucks. The 1987 TIUS population estimates presented in this paper exclude all vehicles identified as a pick-up, van, minivan, utility vehicle, or station wagon on truck chassis. In addition, a vehicle was excluded if the empty combination weight was 6,000 lb or less and the power unit was coded as having only four tires. This should ensure that only light-duty vehicles are excluded from the analysis. However, it is likely that not all light trucks in TIUS were excluded, thus inflating population estimates. Medium-duty straight trucks are the vehicles most likely to be overstated.

To summarize to this point, the difficulty of accurately identifying large trucks in TIUS data probably results in inflated estimates of straight trucks compared with NTTIS. The problem should be less severe for the 1982 file, because it contains a Polk-derived GVWR variable that should be only slightly less accurate than the GVWR determinations employed by NTTIS. The 1987 TIUS straight truck estimates are undoubtedly more affected because that file does not contain a GVWR variable. The GVWR problem is not thought to seriously affect population estimates of tractors in either TIUS file.

NTTIS Time Gap

The third major factor affecting vehicle population estimates between NTTIS and TIUS concerns a time delay in NTTIS between drawing the sample and implementing the survey. The sample was based on the July 1, 1983, R. L. Polk files, but the NTTIS implementation phase was not conducted until January through May of 1985. Vehicles that were junked or scrapped in the interim were removed from the sample, and there was no opportunity to replace them with vehicles that were purchased during that time. This means that NTTIS vehicle counts are low by about a model-year class and a half—those trucks bought during 1984 and the second half of 1983.

In the case of TIUS, the sample is drawn from registrations as of July 1 in a particular year, and survey forms are mailed out over several months of the following year. However, if a vehicle has been junked or scrapped in the meantime, it is still included in the survey. Thus TIUS population estimates refer to the date of the registration lists on which the sample was based, with no loss of cases. Other things being equal, TIUS population estimates should come closer to approximating the entire registered truck population on a given date.

Reconciling Population Estimate Differences

It is possible to adjust NTTIS vehicle count estimates to account for the year and a half of missed model years. Distributions of 1982 and 1987 TIUS vehicle counts by power unit type and model year were examined to see what percent the newest year and a half of model years represent in those two files. Because TIUS samples were based on Polk vehicle lists made halfway through a calendar year, trucks of the newest model year represent about half of a model-year class in TIUS. The next most recent model year should represent a full model-year class. The newest model year and a half of straight trucks represent 4.6 percent of straight trucks in 1982 TIUS and 8.8 percent in 1987 TIUS. It is impossible to say exactly what percent the missed straight trucks in NTTIS represented of the entire straight-truck population when that survey was conducted. The size of model-year classes varies from year to year, as the two TIUS percentages illustrate, because of economic conditions and other factors. However, using the TIUS percentages to estimate a range of missed straight trucks results in an adjusted NTTIS population estimate of 2,291,017 to 2,395,678 vehicles. This is still 5 to 10 percent below the 1982 TIUS estimate and 26 to 29 percent below the 1987 TIUS estimate. Considering that the three surveys were conducted in different years, that it is problematic to identify large trucks in TIUS (especially the 1987 TIUS), and that the adjustment is a rough estimate, the agreement among surveys is not bad.

The newest model year and a half represents 10.5 percent of tractors in 1982 TIUS and 13.1 percent in 1987 TIUS. This results in NTTIS adjusted tractor counts of 1,027,600 to 1,058,755 vehicles. These estimates are 14 to 18 percent above 1982 TIUS tractor counts and between 1 percent below and 2 percent above 1987 TIUS tractor counts. This is also a fairly good agreement, and the adjustment to NTTIS may in fact be higher than is appropriate because of the variation in model-year class sizes.

File Comparisons

Leaving aside the question of absolute vehicle population estimates, NTTIS and TIUS will be compared based on the distribution of several variables describing the large-truck population of each. Both surveys were designed to describe the U.S. registered truck population. Agreement between the two would indicate that they are characterizing the same basic population of vehicles. Although the surveys were conducted in different years, many aspects of the large-truck population change slowly enough that general agreement should be expected if both surveys are representative of the U.S. truck population.

GVWR provides a good basis of comparison between NTTIS and 1982 TIUS because GVWR was included in the original sample data provided by R. L. Polk to both surveys. Most of the other information collected by the two surveys came from respondents and is therefore subject to respondent error. A comparison of the distributions of the national truck population by GVWR from 1982 TIUS and NTTIS is shown for straight trucks in Figure 1 and for tractors in Figure 2. In general, the agreement is good, especially for tractors. The main difference is a somewhat higher proportion of GVWR Class 3 to 5 (10,001 to 19,500 lb) straight trucks in TIUS compared with NTTIS, possibly a result of misclassifications in the Polk-derived GVWRs used in the TIUS file.

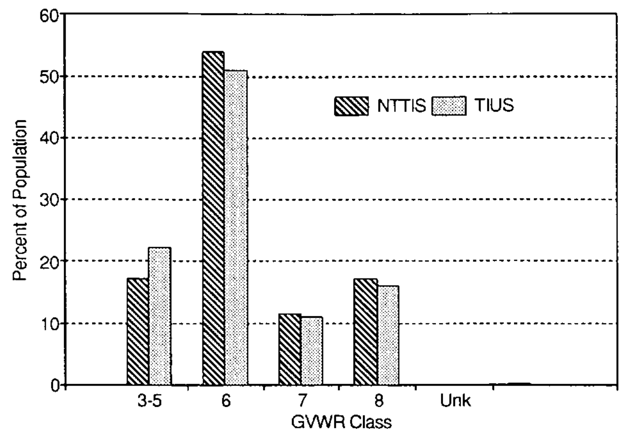


FIGURE 1 Straight trucks by GVWR in NTTIS and 1982 TIUS.

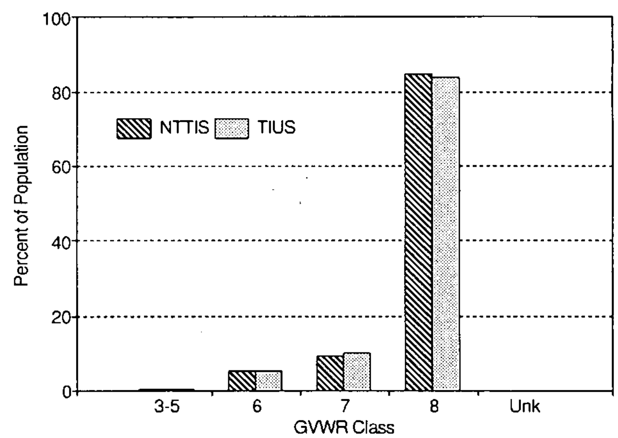


FIGURE 2 Tractors by GVWR in NTTIS and 1982 TIUS.

Compared in Figure 3 is the distribution of cab style for tractors only in NTTIS and 1982 TIUS. This information is obtained from the survey respondent, and the categories used were cabover, short conventional, medium conventional, and long conventional. (Conventional cabs were not subdivided in the 1987 TIUS, so no distribution is included.) The agreement between 1982 TIUS and NTTIS is very good. This is particularly gratifying in view of the lack of a precise definition of what constitutes a short, medium, or long conventional cab.

The last comparison presented here is carrier type for tractors only, shown in Figure 4. Again, this information is supplied by the respondent in both surveys. Carrier types are defined according to whether the company operates interstate or intrastate and whether it is private or for hire. Private carriers operate close to 50 percent of the tractors in both of the TIUS files, and about 53 percent in NTTIS. In NTTIS, a further breakdown of private carriers is made into interstate and intrastate carriers (not shown in Figure 4). Interstate private carriers operate 32.5 percent of all tractors and intrastate 19.9 percent in NTTIS. The remainder of the vehicles in NTTIS and TIUS are for hire in one way or another. For-

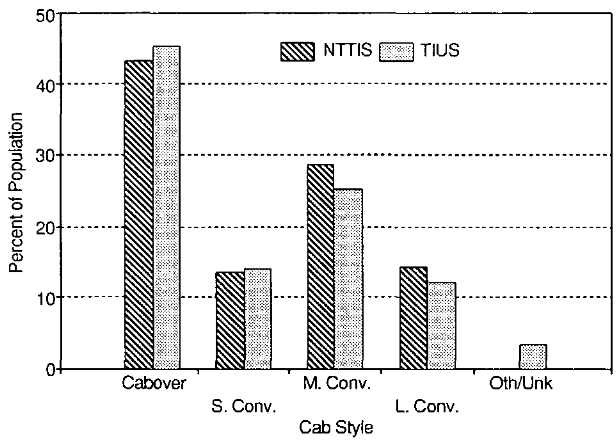


FIGURE 3 Tractors by cab style in NTTIS and 1982 TIUS.

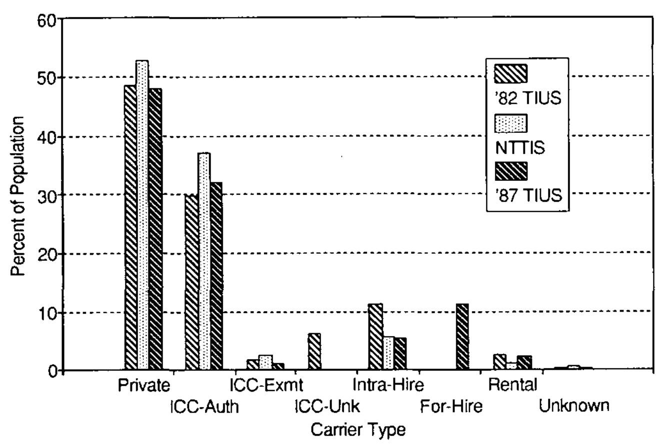


FIGURE 4 Tractors by carrier type in NTTIS and 1982 and 1987 TIUS.

hire vehicles are further subdivided in both NTTIS and TIUS into interstate for-hire, in which case they are subject to Interstate Commerce Commission regulations, and intrastate for-hire, where they are governed by state public service commission regulations. Interstate for-hire vehicles are also separated into authorized carriers—the common and contract carriers—and those hauling exempt commodities. The small group of unknown ICC-regulated carriers in 1982 TIUS are those instances in which respondents did not specify whether they were authorized or exempt carriers. If these cases were distributed between authorized and exempt carriers, it would bring the 1982 TIUS survey into fairly good agreement with NTTIS.

A category of just "for-hire" carriers is included for the 1987 TIUS file. These are cases in which the respondent indicated that the company was for hire but did not specify whether it was subject to ICC regulations. The "for-hire" cases would be distributed among the ICC-authorized, ICC-exempt, and intrastate for-hire categories. This redistribution of cases would probably result in NTTIS having a slightly lower proportion of intrastate for-hire carriers than either TIUS file. NTTIS shows relatively fewer daily rental trucks as well. The owners in both of these categories are usually small carriers and difficult to reach except at night and on weekends. These response problems may be partly responsible for the smaller proportion of trucks operated by intrastate for-hire carriers or in daily rental in NTTIS. Overall, however, the agreement between NTTIS and TIUS on carrier type is quite good.

Truck Travel

Self-Reported Average Annual Mileage Comparisons

NTTIS estimated average annual travel of trucks in three ways: owners' estimates, odometer readings, and mapped mileage from survey calls (*1*). TIUS relies only on estimates from respondents, so NTTIS owner estimates will be used to compare average annual mileage between the two surveys. Both surveys asked owners essentially the same question about how far their truck is driven in a year. Comparisons are based on average annual mileage per vehicle rather than total miles logged by the entire registered large truck population so that the different vehicle population estimates produced by NTTIS and TIUS will not affect the evaluation of mileage estimates.

As shown in Figure 5, the overall agreement in owner-reported average annual travel between the surveys is quite good. The NTTIS straight truck figure is about 18 percent higher than 1982 TIUS and about 13 percent higher than 1987 TIUS. The estimates for tractors are closer, with NTTIS 4 percent higher than 1982 TIUS and 2 percent lower than 1987 TIUS. It is interesting to note that there is a higher degree of correspondence between the files for tractors than for straight trucks. This may be related to the inclusion of some light trucks in the TIUS straight-truck estimates. Light trucks would be expected to travel less in a year, thus lowering the straight truck average.

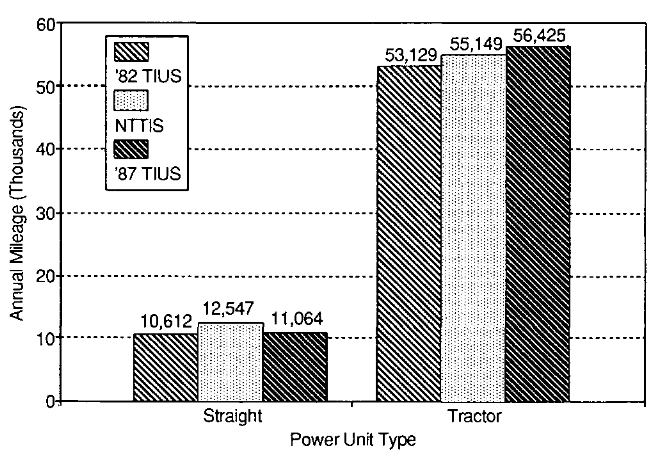


FIGURE 5 Owner-reported average annual mileage by power unit type in NTTIS and 1982 and 1987 TIUS.

Total Annual Mileage

Total mileage estimates by power unit type may also be compared between NTTIS and TIUS. Earlier the degree of undercounting of vehicles in NTTIS because of missed model years was estimated. The corresponding lost travel may be calculated in a similar manner. Straight trucks of the newest model year and a half represent 10.3 percent of the total mileage of 1982 TIUS straight trucks and 15.7 percent of the straight truck mileage in 1987 TIUS. Making the corresponding adjustment to NTTIS raises NTTIS straight-truck mileage from 26,700 million mi to the range of 29,750 to 31,672 million mi (Figure 6). This places NTTIS straight-truck mileage estimates between the 1982 and 1987 TIUS estimates, as would be expected because NTTIS was intermediate in time between the two TIUS surveys. The adjusted NTTIS estimates are 11 to 18 percent above the 1982 TIUS estimate and 11 to 17 percent below the 1987 TIUS estimate.

The newest model year and a half account for 16.3 percent of total tractor mileage in 1982 TIUS and 19.3 percent in 1987 TIUS. This adjustment raises the NTTIS tractor mileage estimate from 49,921 million mi to the range of 59,632 to 61,879 million mi (Figure 7). This places estimated tractor mileage in NTTIS 25 to 30 percent above 1982 TIUS and 2 to 6 percent above 1987 TIUS. The adjusted NTTIS mileage is higher than expected, possibly because the missed model years in NTTIS represented a lower proportion than was calculated using the TIUS files.

Discussion of NTTIS and TIUS

Estimates of national truck population and travel from NTTIS were compared with 1982 and 1987 TIUS. The comparisons covered power unit type, GVWR class, cab style, carrier type, and owner-reported annual mileage. Overall, there is a good correspondence between the two surveys. Some of the differences observed may be because of the different years of registration files from which the samples were drawn, the 18-month period between the sample year and the survey in NTTIS, and the probable classification of some straight trucks with GVWRs below 10,000 lb as Class 3 or higher in TIUS. Aside from these known discrepancies, there is no indication of systematic differences between NTTIS and TIUS.

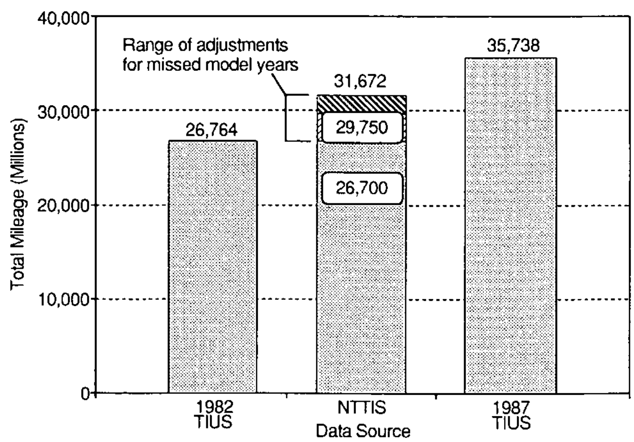


FIGURE 6 Total annual mileage for straight trucks in NTTIS and 1982 and 1987 TIUS.

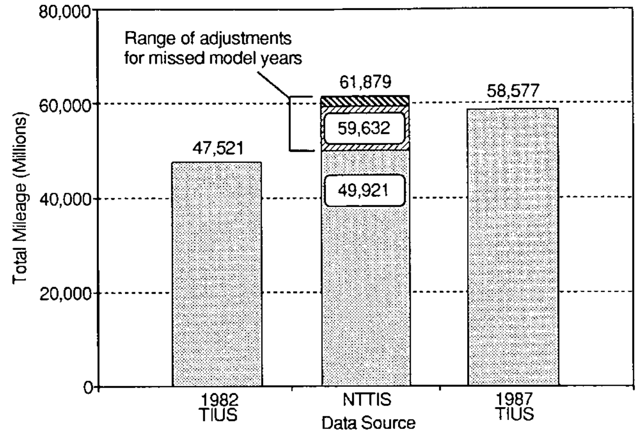


FIGURE 7 Total annual mileage for tractors in NTTIS and 1982 and 1987 TIUS.

COMPARISONS WITH FHWA *HIGHWAY STATISTICS*

Each year the Federal Highway Administration (FHWA) publishes *Highway Statistics*, a tabulation of national transportation statistics based on data submitted by the states. *Highway Statistics* categorizes travel for different classes of vehicles on different types of roads. This section compares national estimates of the number of registered large trucks and their annual mileage from *Highway Statistics* with NTTIS and TIUS estimates.

Data Sources

Highway Statistics categorizes large trucks as single units and combination vehicles. Single units essentially include straight trucks alone, straight trucks hauling utility trailers, and bobtails (tractors without a trailer). Combinations include tractors hauling one or more trailers, as well as straight trucks hauling full trailers. *Highway Statistics* is published annually, and estimates from one year are revised in the following year's edition. The data cited here come from the 1986 and 1988 editions of *Highway Statistics*, Table VM-1 (*2*), representing the revised estimates for the 1985 and 1987 large truck populations respectively. *Highway Statistics* 1985 will be compared with NTTIS, and *Highway Statistics* 1987 will be compared with 1987 TIUS. Numbers for single units were not available for 1982 *Highway Statistics*, so no comparisons will be made with 1982 TIUS.

The *Highway Statistics* data include government-owned vehicles and vehicles registered in Alaska and Hawaii. These vehicles should be excluded for purposes of comparison with NTTIS and TIUS estimates. Because the published *Highway Statistics* data for trucks do not indicate the percentage of government vehicles or the distribution of vehicles by state, estimates were made using other sources of information. The Alaska and Hawaii adjustments for vehicle counts were made based on the state distribution in 1987 TIUS. The Alaska and Hawaii travel estimate adjustments relied on several years of raw and adjusted state-reported mileage figures submitted to FHWA (*3*, *4*). It was more difficult to estimate the percentage of government-owned vehicles because they are not included in TIUS or NTTIS. The vehicle count adjustments for government trucks were made based on an UMTRI data base of large trucks involved in fatal accidents (*5*), and the mileage adjustments took into account figures cited by Mingo (*4*).

The NTTIS vehicle count and mileage estimates used for the comparisons are based on the adjusted figures that account for the missed model year and a half of trucks. The midpoint of the adjusted range was used in each instance. Estimates were produced following *Highway Statistics'* single-unit and combination vehicle classification system. NTTIS mileage figures are based on owner-reported estimates. The TIUS 1987 data were also made consistent with the *Highway Statistics* classification system, but this was slightly more difficult because TIUS produces no estimates for bobtails. Adjustments were made using configuration distributions from NTTIS.

Vehicle Count and Mileage Comparisons

Vehicle count estimates of single-unit trucks are 2,367 million for NTTIS, 3,709 million for 1985 *Highway Statistics* (HS), 3,206 million for 1987 TIUS and 3,668 million for 1987 HS. As noted earlier, the 1987 TIUS straight truck estimate is believed to be too high because of the inadvertent inclusion of light trucks. Given this, it is significant to observe that both HS estimates are even higher than the 1987 TIUS figure. HS estimates more than 14 percent more single-unit trucks than TIUS for 1987.

Vehicle count estimates for combination vehicles are 1,019 million for NTTIS; 1,393 million for 1985 HS; 1,062 for 1987 TIUS; and 1,409 million for 1987 HS. For 1985, HS is about 37 percent higher than NTTIS, and for 1987, HS is 33 percent higher than TIUS. These estimates suggest good agreement between NTTIS and TIUS and a substantial overestimation by *Highway Statistics*.

Total annual mileage estimates are shown in Figures 8 and 9. There is considerable variation in the single-unit travel estimates, with NTTIS showing 29.5 billion mi and 1987 HS estimating 48.3 billion mi of travel (Figure 8). The HS 1985 estimate is 56 percent higher than NTTIS, and the HS 1987 estimate is 37 percent higher than TUIS. The situation is similar for combination travel (Figure 9). HS 1985 estimates 28 percent more mi than NTTIS, whereas HS 1987 estimates 45 percent more travel than TIUS.

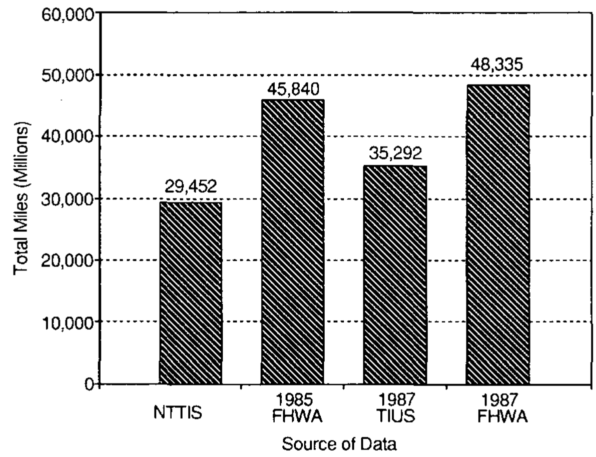


FIGURE 8 Total annual mileage for single-unit vehicles in NTTIS, 1985 and 1987 *Highway Statistics*, and 1987 TIUS.

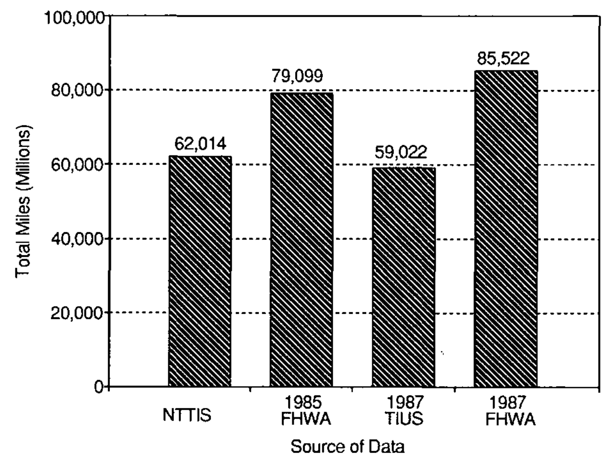


FIGURE 9 Total annual mileage for combination vehicles in NTTIS, 1985 and 1987 *Highway Statistics*, and 1987 TIUS.

Discussion of *Highway Statistics* Estimates

The vehicle count and travel estimates published in *Highway Statistics* are based on data provided by the states. The aggregate statistics are calculated by FHWA using procedures that are intended to provide comparability of values among states. In a recent discussion of *Highway Statistics* large-truck travel estimates, Mingo (*4*) cited several indications that the estimates are too high. Mileage data submitted by states are based on traffic counts of 13 vehicle classes on selected segments of 12 types of roads. Most states use manual and automatic vehicle counting procedures, both of which are problematic. Human error in manual counting often results in the misclassification of vehicle types. With automatic classification, detector deficiencies can result in closely spaced separate vehicles being counted as a single combination vehicle or in the unintended counting of vehicles in adjacent lanes. Because large trucks represent a small proportion of vehicles overall, counting errors can lead to large percentage errors in vehicle class estimates, especially if there is a systematic bias in the miclassifications. Aside from these problems, states do not all employ the same vehicle type classification system. A particular difficulty is straight trucks with trailers, which, depending on state and trailer type, may be classified as either single-unit or combination vehicles.

Another major source of error is that most states count trucks only on weekdays. Generally no correction is made for the fact that truck travel is heavier on weekdays than weekends. Compounding the problem is the fact that counting sites frequently occur on routes with a large volume of heavy trucks.

In addition to these methodological problems, Mingo described other inaccuracies and inconsistencies in state reporting procedures. State estimates in various travel categories have a low level of precision, with mileage figures sometimes reported with only a single significant digit. In most of the states, vehicle-type classifications are entirely omitted for at least some of the road-type breakdowns. Mingo observed many instances of tremendous annual variation in travel estimates within states, including one state that reported an annual increase of more than 500 percent in combination travel.

FHWA attempts to compensate for some of the problems in the state data by adjusting the estimates. For example, a citation on Table VM-1, *1988 Highway Statistics*, indicates that the "stratification of the truck figures is based on the 1982 Truck Inventory and Use Survey (TIUS)." The problem of making these adjustments is compounded because the more recent 1987 TIUS data did not become available until nearly January 1991. The authors cannot evaluate the FHWA adjustment procedures because they have not had the opportunity to review them. Mingo concludes that FHWA's efforts to correct state-reported data contribute to an overestimation of large-truck travel.

The point here is that the *Highway Statistics* figures systematically overestimate large-truck travel. This is a matter of concern because *Highway Statistics* figures are widely used, both in virtually all FHWA studies requiring truck travel data and in many other studies as well. The following example illustrates the relevance of accurate travel information to traffic safety studies. Since 1980 UMTRI has conducted the Trucks Involved in Fatal Accidents (TIFA) survey. The survey combines information from Fatal Accident Reporting System (FARS) cases, Office of Motor Carriers accident reports, and telephone interviews to produce a file of detailed descriptions of all large trucks in the continental United States involved in fatal accidents. In Figure 10 the annual number of fatal involvements of combination vehicles has been plotted for the 7 years from 1982 through 1988 (*5*). The frequency of fatal involvements has remained relatively stable over this period, with a low of 3,376 in 1982 and a high of 3,762 in 1985. On the same graph, the original *Highway Statistics* estimates of the total mileage of combination vehicles for each year have been plotted (*2*). *Highway Statistics* mileage estimates have risen every year. The 90,149 million mi estimated for 1988 represent a nearly 50 percent increase over the 60,310 million mi estimated for 1982. The combination of the substantial increases in estimated travel and the comparatively steady number of fatal involvements results in a sharply declining fatality rate. This is also plotted in Figure 10, against the *y*-axis on the right edge of the graph. According to the *Highway Statistics* numbers, the fatal involvement rate of combination vehicles/100 million mi of travel has declined from 5.60 in 1982 to 4.12 in 1987, a drop of 26 percent.

Although such a dramatic decrease in the fatal involvement rate would be encouraging news, it is possible that much of this trend is an artifact of systematic error in the *Highway Statistics* travel estimates. It is reasonable to believe that large-truck travel has increased from year to year, with the overall expansion of the economy. However, TIUS estimates only a

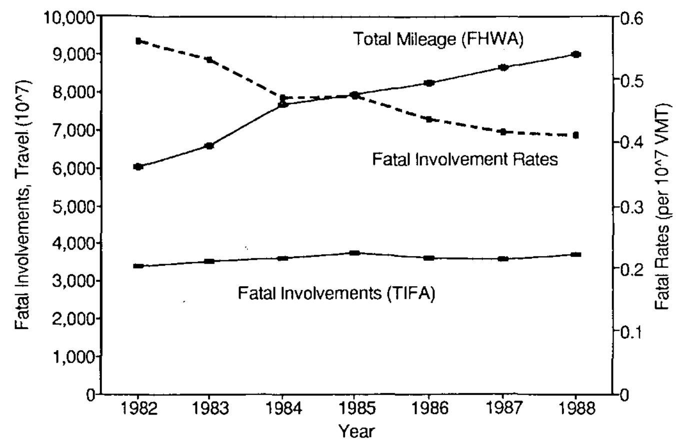


FIGURE 10 Fatal involvement rate of combination vehicles, 1982–1988.

23 percent rise in tractor travel from 1982 to 1987, whereas *Highway Statistics* estimates a 43 percent increase in combination vehicle travel during the same time span. Furthermore, *Highway Statistics'* 1982 figure was only 27 percent higher than the 1982 TIUS estimate, whereas in 1987 the *Highway Statistics* figure was 47 percent above TIUS. This suggests that cumulative error in the *Highway Statistics* large-truck travel estimates increases the amount of overestimation over time. If the *Highway Statistics* mileage figures are too high, then fatal involvement rates based on those figures will be too low.

CONCLUSIONS

Accuracy of large-truck travel estimates is clearly an important issue. Evaluating the safety of particular classes of vehicles requires information on both the number of accidents they experience and how many miles they accumulate, so that accident rates per mile of travel may be calculated and compared with other kinds of vehicles. Travel estimates that are too high will produce accident rates that are too low. Compared in this paper are large-truck travel estimates from three sources. The comparisons are not as straightforward as desired because of the different times the data were collected and the different methodologies used by each source. However, the overall conclusion is that estimates produced by TIUS and NTTIS show much closer agreement to each other than either survey does to estimates published in *Highway Statistics*. Ideally, more nationally representative surveys of large-truck travel will be conducted in the coming years so that their results may be included in similar travel comparisons. With more independent studies, the accuracy of *Highway Statistics* estimates can be better evaluated.

ACKNOWLEDGMENT

This analysis of the National Truck Trip Information Survey was supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.

REFERENCES

1. D. L. Massie, K. L. Campbell, and D. F. Blower. Large-Truck Travel Estimates from the National Truck Trip Information Survey. In *Transportation Research Record 1407*, TRB, National Research Council, Washington, D.C., 1993 (in production).
2. *Highway Statistics* (1982–1988 editions). Office of Highway Information, Federal Highway Administration, U.S. Department of Transportation, 1983–1989.
3. R. D. Mingo. *Safety of Multi-Unit Combination Vehicles*. Intermodal Policy Division, Association of American Railroads, 1990.
4. R. D. Mingo. *Evaluation of FHWA's Vehicle Miles of Travel Estimates for Heavy Vehicles*. Intermodal Policy Division, Association of American Railroads, 1991.
5. D. F. Blower. *Trucks Involved in Fatal Accidents, 1980–1988, by Power Unit Type*. Report UMTRI-91-36. University of Michigan Transportation Research Institute, Ann Arbor, 1991.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.