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,

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

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

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

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

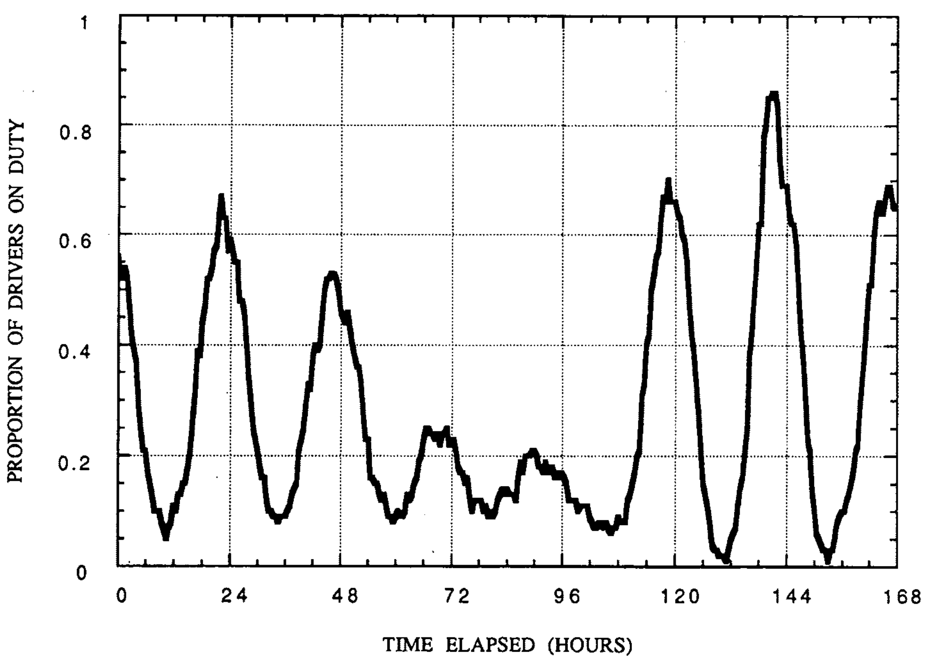


FIGURE 1 Driving Pattern 1.

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.

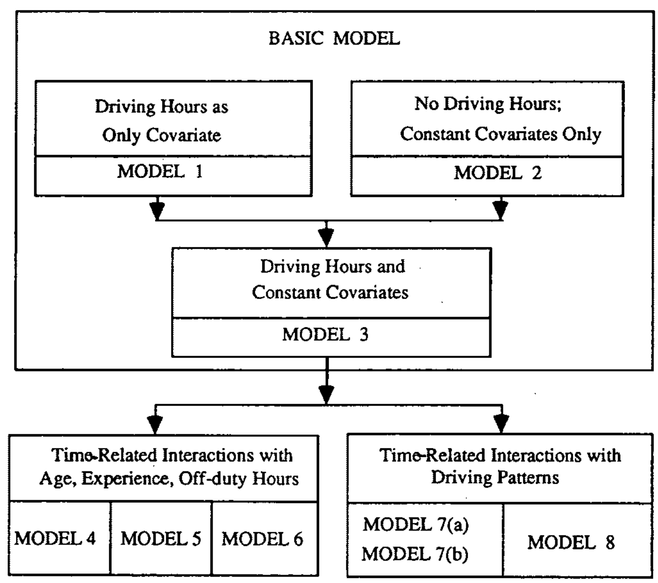
The goodness-of-fit of a model to the data can be qualitatively assessed by plotting model values as a function of driving time against the product limit estimate of the data (10,16). The survival function is denoted as

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

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 *		2.4878 *		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)	0.5811	(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)	0.4879	(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

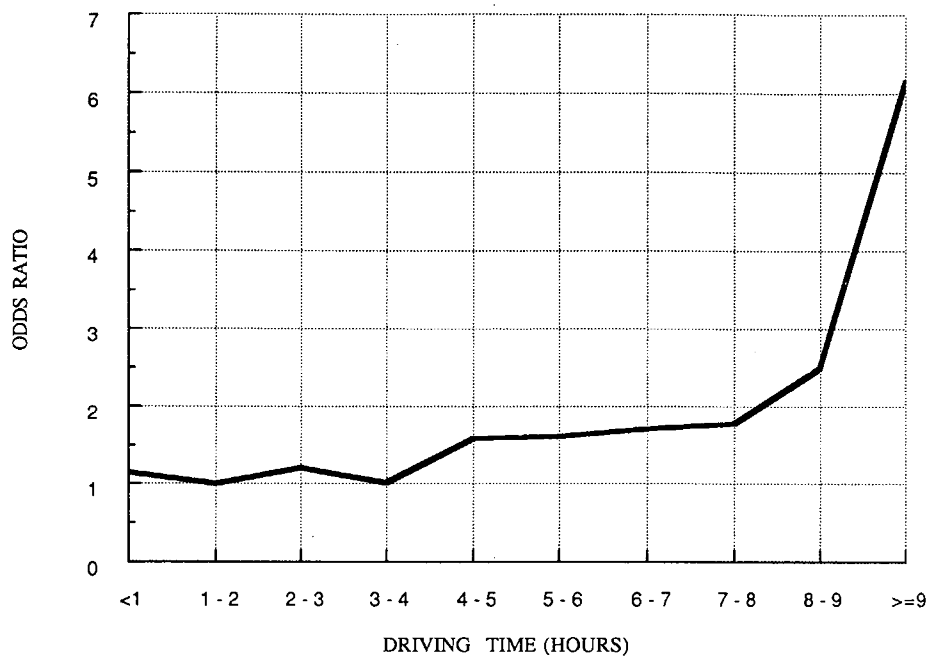


FIGURE 2 Odds ratio of driving hours.

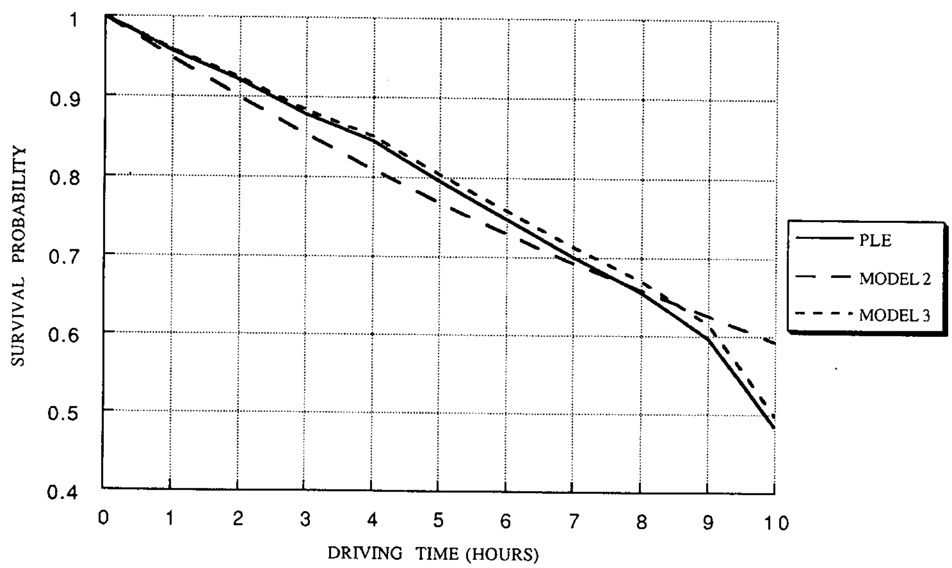


FIGURE 3 Survival curve.

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-

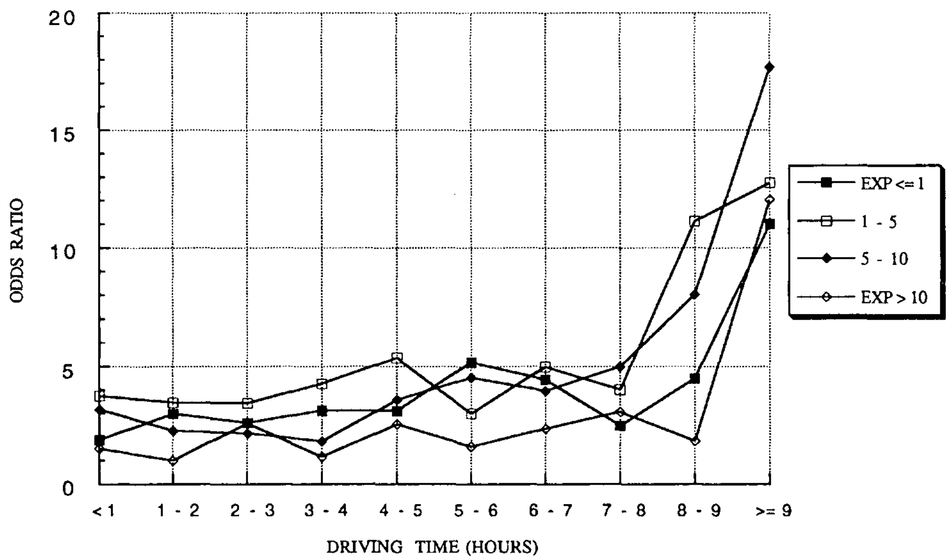


FIGURE 4 Odds ratio of driving experience over time.

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.