Time of Day Models of Motor Carrier Accident Risk

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) until that time. Three logistic regression models are estimated, which include time main effects (the driving time), time-independent effects (experience), time-dependent effects (time of day), 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 4th hr, by approximately 50 percent or more, until the 7th hr. The 8th and 9th hr show a further increase, approximately 80 and 130 percent higher than the first 4 hr. Drivers with more than 10 years of driving experience retain a consistently low accident risk; all other categories of driving experience have a significantly higher risk. Daytime driving, particularly at the noon time (10:00 a.m. to 12:00 noon), results in a significantly lower risk of an accident. Drivers at one time of day (4:00 to 6:00 p.m.) have an accident risk about 60 percent higher than those driving during the baseline; drivers during the other three significant times of day also experience accident risks about 40 percent higher than drivers during the baseline. All three times of day involve night or dawn driving; two are associated with circadian rhythms. Rest breaks, particularly those taken before the 6th or 7th hr of driving, appear to lower accident risk significantly for many times of day.

Motor carrier safety has been an area of active study throughout the 1980s and the early 1990s. Of the factors generally considered in safety studies (i.e., driver, vehicle, roadway, and environment), particular attention has been paid to driver-related factors. One major study concluded that 65 percent of accidents may be attributable to human errors (*1*).

Driving fatigue is believed to have a particularly powerful effect on commercial vehicle drivers, representing one of the primary human factors. Fatigue significantly increases driving errors and decreases driver alertness. Two additional studies using restricted data bases have found more than 30 percent of heavy truck crashes may result from driving fatigue (*2,3*). Nevertheless, fatigue is a sufficiently vague concept in that it has not been precisely defined and measured (*4*), a fact that presents difficulties in applying fatigue concepts in accident models. Several studies have described factors associated with either physiological or psychological components of fatigue (*4–7*).

Driving hours, for one origin-to-destination trip or over several trips and multiple days, is often an important element of fatigue. Several studies have considered the appropriateness of government-regulated limits on driving hours. These studies seek to identify hours that pose higher accident risk and policy changes that could result in reduced accident risk (*8–14*). Although it may seem straightforward to account for the influence of driving hours on fatigue, there are many subtleties to be considered. Among those already studied in the literature are the effects of the following: off-duty hours immediately before a trip and multiday driving (*13,14*); heat, noise, and vibration (*15*); cargo loading and unloading (*16*); patterns of rest in sleeper berths (*17*); and alcohol and drugs (*3,18*).

This research attempts to contribute to this literature by identifying the effects of time of day on accident occurrence. Circadian rhythms, which are changes in body function following an approximate 24-hr cycle, are of particular importance. Although circadian rhythms vary somewhat from person to person, the most common pattern is one with a physiological low around 4:00 to 6:00 a.m., representing a time of particular risk to drivers. This represents a substantial societal risk as a significant amount of truck travel occurs at night.

Several relevant studies have focused on the relationship between motor carrier accidents and time of day. Harris and Mackie (*8*) concluded that the lowest levels of alertness occur for most drivers between midnight and 8:00 a.m. Several additional studies also have found elevated involvement or accident risk in this same time interval, suggesting a circadian effect (*9,17,19,20*). Interestingly, Hamelin (*10*) indicated that accident involvement rates generally increase throughout the day from a low point around 4:00 to 6:00 a.m. to a high point from midnight to 2:00 a.m. There is also a sharp peak in risk around noon. Another study of automobile drivers found that an additional period of decreased alertness occurs in the mid-afternoon (*21*). This research aims at a more explicit quantification of the effect of time of day on motor carrier accident risk.

OBJECTIVES

There is a need to develop quantitative methods to analyze the effect of time of day on accident risk. In particular, it is important to consider whether the circadian effect plays a major role in motor carrier accident risk. One objective of this study is to use time-dependent logistic regression to formulate a quantitative model that explicitly includes time of day along with other covariates. The second objective is to test the model using data from actual trucking company operations and to compare the results with those in the extant literature.

LOGISTIC REGRESSION MODEL

A general formulation for the time-dependent logistic regression model is as follows:

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

Institute of Transportation Studies, Department of Civil and Environmental Engineering, University of California—Davis, Davis, Calif. 95616.

in which Y_{ti} is a response variable representing the occurrence ($Y_{ti} = 1$) or nonoccurrence ($Y_{ti} = 0$) of the event for individual i during the time interval t. X_i is an univariate or multivariate attribute vector for this individual, and $g(X_i, t, \beta)$ denotes some arbitrary function of X_i and a parameter vector β that will be estimated (*22–25*). In accident analysis, the conditional probability expressed in Equation 1 is the probability of an accident at time interval t, given survival (i.e., no accident) until that time; in other words, the model accounts for the survival effect (*14*). In this case, 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.

The comparable conditional probability of surviving is defined as

$$Q_{it} = 1 - P_{it} \tag{2}$$

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

$$g(X_i, t, \beta) = \sum_{j=0}^{r} \beta_j X_{ji} \tag{3}$$

The X_{ji} ($j = 0, \ldots, r$) are the values of the r covariates for the driver i. The value of $X_{0i} = 1$ so that β_0 represents an intercept parameter in the regression model.

The full likelihood for the n drivers can be represented by the following:

$$L = \prod_{i=1}^{n} \left(\frac{P_{i\,t_i}}{Q_{i\,t_i}} \right)^{Z_i} \prod_{t'_i \le t_i} Q_{i\,t'_i} \tag{4}$$

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.

Equation 3 can be broken down into the following components:

$$g(X_i, t, \beta) = \sum_{j=0}^{r} \beta_j X_{ji} + \sum_{k=1}^{T-1} \beta_{r+k} X^{*}_{ki} + \sum_{n=1}^{s-1} \beta_{r+(T-1)+n} X_{ni}(t_i) \tag{5}$$

The first term of the right-hand side of Equation 5 represents time-independent covariates, the effects of which are assumed to be independent of time. The second term represents the time main effect (in this application, driving time), and X^{*}_{ki} represents the kth time interval for driving time. A trip with a length of k time intervals would be represented by a series of indicator variables with $X^{*}_{ki} = 1$. The last term represents the time-dependent covariate (in this application, time of day). The parameters $\beta_{r+(T-1)+n}$ are a series of coefficients associated with the s intervals used as categories for the time-dependent covariate (in this case, 11 categories of time of day). A similar model formulation was used elsewhere (*14*); Equation 5 represents an extension of the earlier model in that it includes time-dependent covariates.

To include the survival effect in the time-dependent logistic regression model correctly, several methods to treat time dependent covariates have been proposed. One approach (*26*) specifies a series of covariates to represent each time-dependent risk factor for each time interval. A nice feature of this method is that it suggests approaches to incorporate change in the underlying risk of an event over time and the prior history of an individual. However, the model in this general form could contain so many parameters to be estimated that it might be difficult to interpret.

Another approach (*27*) uses a parsimonious model to reduce the dimensionality of the model and to improve its interpretability. A duplication method is developed to overcome the assumption of standard logistic regression that restricts each individual to only one ultimate outcome. As an example of the method, consider a driver with an accident in the third time interval. Three records will be generated for this case. For the first two records, the values of the response variable are both 0 (non-accident); the value of the response variable will be 1 for the third record. For a driver who successfully completes a trip through the third interval, three records will also be generated; the values of the response variable for all three records are 0. The values of the vector of time-independent covariates for this individual will be the same in each of the three records, whereas the values of time varying covariates will depend on the related time interval. This approach is based on the following three important assumptions:

1. The underlying risk of the events in each time interval is assumed the same in this model (e.g., the risk in the first driving hour is the same as that in the 9th hr).
2. Closely related to the first assumption is that risk factors and outcome of interest are independent of time; that is, for a particular time of day (e.g., 8:00 to 10:00 a.m.) the accident risk for the 1st hr of driving is the same as that during the 9th hr of driving.
3. Only the current status of the risk factor is associated with the outcome of the event, prior history is considered unimportant.

In this research, the approach (*27*) to treat the repeated measurements will be followed, but time will be treated as categorical in the model to reflect underlying risk. This relaxes the first assumption of Cupples's model. The second assumption is relaxed by including in the model interaction terms to address the potential association between driving hours and time of day.

DATA AND VARIABLE DESCRIPTION

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 takes reasonable steps to adhere to Department of Transportation service hour regulations, most drivers in the study can be considered to comply with existing limits. The data include accidents and non-accidents from the company's national over-the-road operations.

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 being struck by a truck, but no report was filed or verified by the carrier). The severity ranges from minor fender benders to accidents with fatalities. A non-accident is defined as "the case in which a driver successfully completes the designed trip." This is generally called "censoring" data because the accident cannot happen after the designed trip is finished.

The time-dependent logistic regression is developed using variables that include the experience of the driver with the firm, the consecutive hours of driving on the trip in question, and the time of day. The consecutive hours of driving are the actual driving time based on the designed trip of interest that restricts the maximum driving hours limits until the accident occurs or the trip is completed. The off-duty time (short breaks) and the time on-duty without driving (intermediate terminal) during this trip are then excluded.

It is possible for a driver to make several short stops either because of feeling tired or because of an intermediate terminal. These stops do not end the trip in terms of the measured time duration t_i in Equation 4; time will accumulate after the stops until an accident or completion of the trip at the destination terminal. The driver is not given the option to terminate a trip and simply stop to sleep. The truck must reach a destination at a particular time. Therefore, the driving time is either the time to an accident or the time to censoring, each of which is independent.

A problem arises because of the need to code time of day as a series of dummy variables to account for possible non-linearities. To keep the estimated number of parameters to a reasonable scale, 2-hr time periods were chosen for time of day. In this case, the first interval is midnight to 2:00 a.m.; the twelfth is 10:00 p.m. to midnight. Given that driving time data are recorded at the level of 15 min, raw data on driving time must be converted to a series of more aggregate categorical variables of 2-hr duration. Difficulties arise because drivers take rest breaks and are off duty for some time during a typical day. When these rest breaks and off-duty times occur within a 2-hr time category, it is necessary to assign the driver as either driving or not driving for that unit of time.

The rules that determine the coding of the time of day variable are as follows

- If the driver is driving for an entire time of day represented by the variable, then the driver is coded as driving during that time of day.
- If a driver's driving time crosses more than one time of day period (for example, driving from 1:45 a.m. to 2:45 a.m.), then the most proportional time of day will be coded (in this example from 2:00 a.m. to 4:00 a.m. as driving; from midnight to 2:00 a.m. as not driving).
- If a particular driving time bisects two time-of-day periods exactly, the latter time of day is coded as driving.

In this research the total number of observations used for modeling is 1924 cases, of which 694 are accidents and 1230 are non-accidents. Accidents are deliberately oversampled relative to their actual occurrence to handle the data more efficiently. Although the sampling is a type of case-control method typically used in a retrospective study, the likelihood function in Equation 4 developed for prospective studies can still be applied because the logistic regression is adopted in this research (*28*).

EMPIRICAL RESULTS

Overview of Modeling

An overview of the time-dependent logistic regression models developed in this research is shown in Figure 1. Model 1 is developed to assess the underlying hazard of driving time only. A time-independent covariate, driving experience, and a time-varying covariate, time of day, are added and estimated in Model 2. A series of models is developed to study interactions between time of day and driving time. A separate model is developed with Model 2 and interaction terms with each time of day separately with all nine categories of driving hours. The significant variables are entered into one model and a stepwise deletion procedure used to arrive at the model shown as Model 3 (*14*). The statistical software BMDP is used to estimate coefficients and derive appropriate statistics concerning model fit.

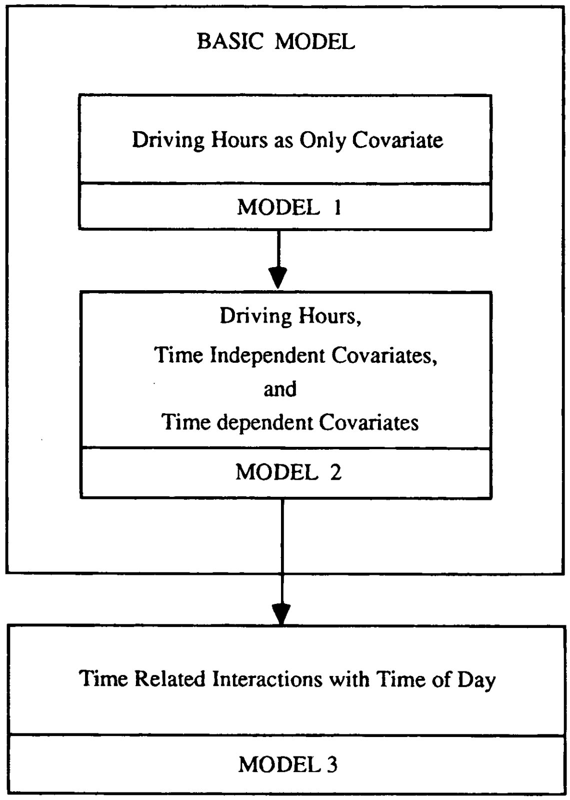

FIGURE 1 Modeling structure.

Several tests are conducted to assess the significance of variables and models. First, a likelihood ratio test for inclusion or exclusion of a variable as a whole is used as an exploratory test of variable significance (e.g., inclusion of all categories of experience). Second, t-statistics are reported for each category of each variable.

The goodness of fit of a time-dependent logistic regression model to the data can be qualitatively assessed by plotting model values as a function of driving time against the product limit estimator (PLE) of the data (*23,24*). The survival function for the logistic regression is denoted as follows:

$$S(t) = \prod_{t' \leq t} Q_{it'} \tag{6}$$

and the survival function for the product limit estimator is

$$S(t) = \prod_{t' \leq t} (N_{t'} - D_{t'})/N_{t'} \tag{7}$$

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 time interval t'. This goodness-of-fit measure has been used elsewhere (*14*).

Basic Models

A basic model that includes only driving hours is shown as Model 1 in Table 1. The model implicitly assumes that the probability of an accident is entirely determined by the driving time and is unaffected by other driver attributes. Model 1 is constructed so that there is a constant hazard within each hour and varying hazards 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 accidents 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. Accident risk is insignificantly different in the first 4 hr but rises steadily thereafter to a maximum in the 10th hr.

Model 2 shows the results of combining Model 1 with driving experience and time of day. The likelihood ratio test between Model 2 and Model 1 is significant beyond $\alpha = 0.05$, which leads to a

TABLE 1 Model Estimates and Statistics

NO	COVARIATES	MODEL 1	MODEL 2	MODEL 3
1	CONSTANT	-3.2780 *	-3.9603 *	-5.3635 *
	EXPERIENCE (year)			
2	<= 1		0.5658 *	0.5553 *
3	1 - 5		0.8210 *	0.8087 *
4	5 - 10		0.5929 *	0.5852 *
5	> 10**			
	TIME OF DAY			
6	0:00 - 2:00		0.3318 #	1.7938 *
7	2:00 - 4:00		0.0407	1.3996
8	4:00 - 6:00		0.2798	1.7563 *
9	6:00 - 8:00		0.3669 #	1.7277 *
10	8:00 - 10:00		0.2452	1.7509 *
11	10:00 - 12:00**			
12	12:00 - 14:00		0.1369	1.7179 *
13	14:00 - 16:00		0.0958	1.4638
14	16:00 - 18:00		0.4920 *	2.0918 *
15	18:00 - 20:00		0.2356	1.7032 *
16	20:00 - 22:00		0.3399 #	1.7293 *
17	22:00 - 24:00		0.0444	1.5051
	DRIVING HOURS			
18	1st HOUR (<1)	0.1404	0.1325	1.5128
19	2nd HOUR (1 - 2)**			
20	3rd HOUR (2 - 3)	0.1835	0.1903	1.5759
21	4th HOUR (3 - 4)	0.0040	0.0143	1.4655
22	5th HOUR (4 - 5)	0.4481 *	0.4673 *	1.8532 *
23	6th HOUR (5 - 6)	0.4628 *	0.4872 *	2.1375 *
24	7th HOUR (6 - 7)	0.5133 *	0.5290 *	2.1183 *
25	8th HOUR (7 - 8)	0.5392 *	0.5670 *	1.9501 *
26	9th HOUR (8 - 9)	0.8625 *	0.9119 *	2.3669 *
27	10th HOUR (> = 9)	1.8377 *	1.8200 *	3.4343 *
	INTERACTIONS			
28	(6) & (23)			-2.2060 *
29	(8) & (24)			-2.7526 *
30	(10) & (21)			-3.0946 *
31	(10) & (27)			-2.6086 *
32	(12) & (23)			-2.4369 *
33	(12) & (24)			-2.4721 *
34	(14) & (23)			-2.8250 *
35	(14) & (26)			-2.9784 *
36	(14) & (27)			-2.7428 *
37	(15) & (23)			-2.4132 *
38	(17) & (24)			-3.1159 *
39	OTHERS			-1.4307
	LOG-LIKELIHOOD VALUE	-2698.74121	-2663.0332	-2641.15161
	LIKELIHOOD RATIO TEST		71.41602 (v.s. MODEL 1)	43.76318 (v.s. MODEL 2)
	DEGREE OF FREEDOM		14	12
	CHI-SQUARE (0.95)		23.685	21.026

\# t STATISTICS SIGNIFICANT @ α=0.10

* t STATISTICS SIGNIFICANT @ α=0.05

** REFERENCED CATEGORY

rejection of the hypothesis of driving time as the only covariate. Time of day alone, without experience, failed to reject the null hypothesis of no effect as a whole.

Parameter values for driving hours in Model 2 are virtually identical to Model 1. The baseline hazard fluctuates from the 1st hr to the 4th hr with no significant difference then increases significantly until the last hr.

Drivers with experience of more than 10 years have the lowest accident risk (baseline category). The accident 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 than for the baseline). The estimated risk increase for drivers with less than or equal to 1 year experience and those with 5 to 10 years of experience is nearly equal (about 1.7 times higher than for the baseline category).

Concerning time of day, drivers in the time between 10:00 a.m. and noon had the lowest risk, so it was defined as the baseline. The accident risk of driving during 4:00 to 6:00 p.m. is significantly higher than that of the baseline, beyond $\alpha = 0.05$. This highest accident risk may result from a combination of two effects: 4:00 to 6:00 p.m. is the evening rush hour in most major cities, increasing accident risk because of the likelihood of a collision with another vehicle; a second effect could be an association with reduced alertness because of a low circadian period for some drivers (*21*). The accident risks from midnight to 2:00 a.m., 6:00 to 8:00 a.m., and 8:00 to 10:00 p.m. are also significantly higher than during the baseline (but at $\alpha = 0.10$). Two of them involve night driving; the other involves part of the dawn period.

Inclusion of Interaction Terms

The modeling of interaction terms between time of day and driving hours is summarized as Model 3 in Table 1. The objective of testing this set of variables is to determine whether certain times of day are particularly risky (or safe) for driving hours of a particular duration. This is an examination of the effect of two time-related covariates. The addition of time-related interactions results in Model 3 having a significantly improved goodness-of-fit compared to Model 2. Figure 2 indicates little difference between the two models in a comparison of their fit to the product limit estimator of Equation 7. The fit appears good.

Consistent with the previous model, the three categories of driving experience in Model 3 have significant positive parameters, and they are of virtually the same magnitude as in Model 2. The parameters for driving hours are similar to Model 2, but the magnitudes change because of the time-related interactions.

All the significant interactions result in the reduction of accident risk for a specific time of day over time. When interaction terms are added, four of the times of day that were indifferentiable from the baseline became significantly higher in risk from the baseline. This also happened for all three of the marginally significant times of day. On the basis of these results, there is no question that time of day and driver hours interact. The interactions thus allow differentiation of times of day of constant elevated risk from those whose risk varies with driving time.

Nevertheless, some times of day have risks no different from the baseline, specifically 2:00 to 4:00 a.m. and 2:00 to 4:00 p.m. The

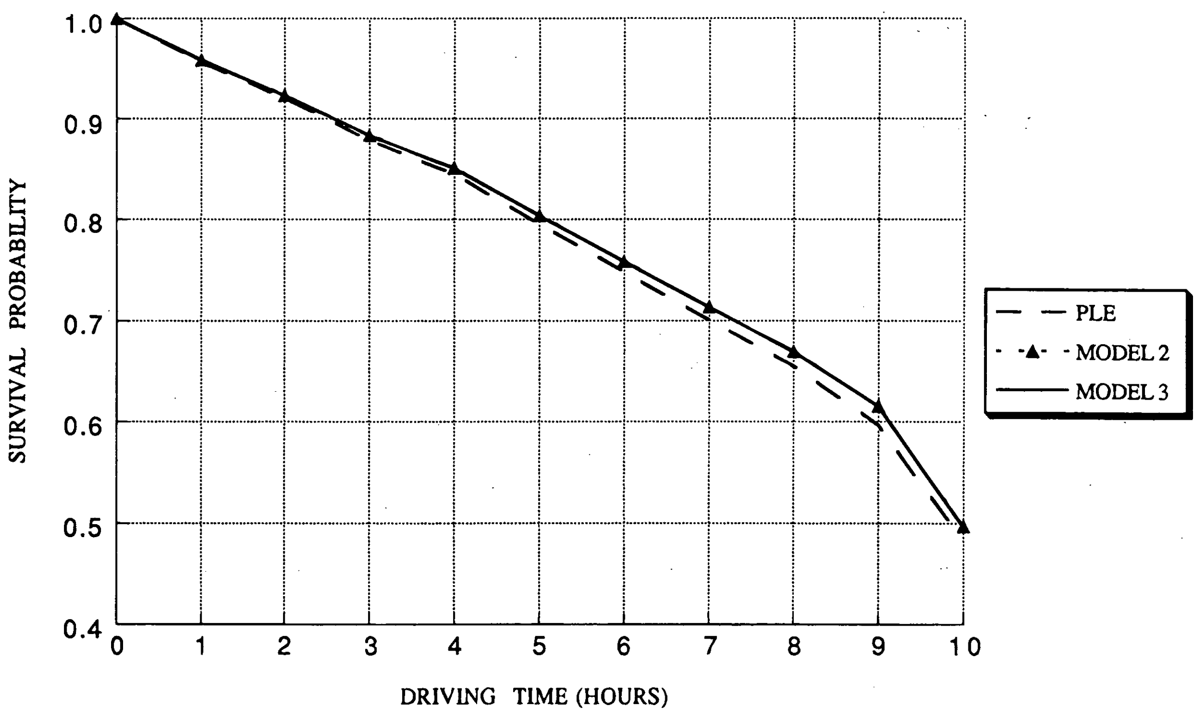

FIGURE 2 Survival curve for model goodness of fit.

TABLE 2 Survival Model Estimates and Statistics

NO	COVARIATES	COEFFICIENT	T VALUE	EXP (COEFFICIENT)
	AGE			
1	<= 39	0.1714	1.5897	1.1869
2	39 - 46**			
3	46 - 53	0.0477	0.4369	1.0488
4	> 53	0.1133	1.0273	1.1199
	EXPERIENCE (year)			
5	<= 2	0.5873	5.0676 *	1.7992
6	2 - 6	0.6428	5.4792 *	1.9018
7	6 - 10	0.5988	5.5288 *	1.8200
8	> 10**			
	DRIVING PATTERN			
9	1	0.1221	0.7881	1.1299
10	2	-0.1995	-1.1266	0.8192
11	3	0.1673	1.0869	1.1821
12	4	-0.0516	-0.3078	0.9497
13	5**			
14	6	0.0511	0.3067	1.0525
15	7	0.1418	0.9021	1.1524
16	8	0.0539	0.3421	1.0554
17	9	0.1068	0.6531	1.1127
	OFF-DUTY HOURS			
18	<= 10.5**			
19	10.5 - 13.75	-0.1311	-1.2004	0.8772
20	13.75 - 25.75	-0.1411	-1.3239	0.8684
21	> 25.75	-0.2072	-1.8946	0.8129
	TIME OF REST (hour)			
22	no rest break**			
23	<= 2	-0.0574	-0.5760	0.9442
24	2 - 4	-0.2015	-2.0307 *	0.8175
25	4 - 6	-0.3747	-3.1971 *	0.6875
26	6 - 8	-0.0383	-0.2343	0.9624
27	> 8	-0.9327	-1.2870	0.3935
	LOG-LIKELIHOOD VALUE	-5066.3731		
	GLOBAL CHI-SQUARE	81.36		
	DEGREE OF FREEDOM	22		
	P-VALUE	0		

* t STATISTICS SIGNIFICANT @ $\alpha = 0.05$

** REFERENCED CATEGORY

time period from 10:00 p.m. to midnight also has indistinguishable risk from the baseline except for a significant and negative interaction with the seventh driving hr. These times of day represent periods of particularly low risk, and they are, with one exception, independent of driving time. Other time periods with significant interactions may have lower risk for some driving times.

The prevalence among the interaction terms of significant interactions with the sixth and seventh driving hours is surprising. On the basis of the literature, there is no a priori expectation for the observation of this systematic risk reduction. Additional modeling of this data set using survival models (*29*) helps to interpret this result further.

Table 2 is the output of a survival model estimation. The model coefficients can be interpreted similarly to a linear regression model. In this case, positive coefficients imply increased risk of an accident, negative coefficients a reduced risk. Age, experience, and off-duty hours before the trip of interest are all listed as categorical variables. The time of day of multiday driving is characterized by a driving pattern number that is the output of a cluster analysis (*13*). Of particular importance to this discussion is the set of "time of rest" categorical variables, which are used to depict the taking of a rest break during a particular driving hour. Notice that rest breaks during driving hours 2 to 4 and 4 to 6 significantly lower accident risk. It appears that the interaction terms in our logistic regression model are picking up this rest break effect. The survival model is presented here strictly to clarify the interpretation of the logistic regression interaction terms. The theory of the survival model is thus not important in this context. The consistency of the effects observed is important.

SUMMARY AND RECOMMENDATIONS

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 main effects, time-independent covariates, time-dependent covariates, and interaction terms. The model examines accident risk 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) until that time interval. Individual accidents are statistically compared with a random sample of individual non-accident trips by estimating a logistic regression model with two outcomes: an accident or non-accident. Covariates tested in the model include consecutive driving time, driver experience, and time of day.

Three logistic regression models are estimated, which include main time effects (driving time), time-independent effects (driving experience), time-dependent effects (time of day), 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 4th hr, by approximately 50 percent or more, until the 7th hr. The 8th and 9th hr show a further increase, approximately 80 and 130 percent higher than the first 4 hr. These results are generally consistent with those of Harris and Mackie (*8*).

Drivers with more than 10 years driving experience retain a consistently low accident risk; all other categories of driving experience have a significantly higher risk than this group.

Time of day had an effect on subsequent accident risk, but the effect was not as strong as for driving experience or driving hours. Daytime driving, particularly at noon (10:00 a.m. to 12:00 p.m.), results in a significantly lower risk of an accident. Driving from 4:00 to 6:00 p.m. has an accident risk about 60 percent higher than the baseline; drivers during the other three significant times of day also have accident risks about 40 percent higher than those during the baseline. These three involve night or dawn driving; two of them are associated with circadian rhythms.

When interactions were included, the accident risk for some times of day decrease. Particularly, most of the significant interactions fall in the sixth and seventh driving hours. Rest breaks appear to be associated generally with these risk reductions.

Time-dependent covariates play a key role in accident analysis. However, the shortage of time-varying data makes it difficult for a researcher to consider further accident analysis and solutions. As mentioned earlier, high traffic volume could be one of the reasons for the highest accident risk occurring between 4:00 and 6:00 p.m. The inclusion of road class (e.g., rural Interstate, urban local), which is a kind of time-varying risk factor, could greatly improve understanding of time-related effects. The collection of this additional time-dependent data becomes an important task in future research.

The joint study of time of day and driving time is complicated because driving time intervals could cross more than one time of day. Although some rules have been provided in this research, the approach is still rough and could result in some loss of information and bias in estimation. A more advanced approach is needed to treat the coding of time of day precisely and completely.

In this research, there is an important assumption that the prior history of an individual does not influence the outcome. Cupples et al. (*27*) used the slope of a risk factor over time to represent the effect of past history on an outcome. Time of day cannot be treated in this way because it is a categorical variable. The inclusion of prior history as a time-dependent covariate, while keeping the model parsimonious, is an important topic of future research.

ACKNOWLEDGMENT

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

REFERENCES

1. *Gearing Up for Safety: Motor Carrier Safety in a Competitive Environment.* Office of Technology Assessment, U.S. Congress, 1988.
2. *A Report on the Determination and Evaluation of the Role of Fatigue in Heavy Truck Accidents.* Transportation Research and Marketing, AAA Foundation for Traffic Safety, Falls Church, Va., 1985.
3. *Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes.* Vol. 1, Report NTSB/SS-90/01, National Transportation Safety Board, Washington, D.C., 1990.
4. Haworth, N. L., T. J. Triggs, and E. M. Grey. *Driver Fatigue: Concepts, Measurement and Crash Countermeasures.* Department of Transport and Communications, Federal Office of Road Safety, Canberra, Australia, 1988.
5. McDonald, N. *Fatigue, Safety and the Truck Driver.* Taylor and Francis, London, England, 1984.
6. Brown, I. D. Fatigue. In *Medical Aspects of Fitness to Drive.* The Medical Commission on Accident Prevention, London, England, 1985.
7. Sussman, E. D., H. Bishop, B. Madnick, and R. Walter. Driver Inattention and Highway Safety. In *Transportation Research Record 1047,* TRB, National Research Council, Washington, D.C., 1985, pp. 40–48.
8. Harris, W., and R. R. Mackie. *A Study of the Relationships Among Fatigue, Hours of Service, and Safety of Operations of Truck and Bus Drivers.* Report BMCS-RD-71-2. Human Factors Research, Inc., Goleta, Calif., 1972.
9. Hackman, K. D., E. Larson, and A. Shinder. *Analysis of Accident Data and Hours of Service of Interstate Commercial Motor Vehicle Drivers.* Genasys Corp., Rockville, Md., 1978.
10. Hamelin, P. Lorry Drivers' Time Habits in Work and Their Involvement in Traffic Accidents. *Ergonomics,* Vol. 30, No. 9, 1987, pp. 1303–1333.
11. Jones, I. S., and H. S. Stein. *Effect of Driver Hours of Service on Tractor-trailer Crash Involvement.* Insurance Institute for Highway Safety, Arlington, Va., 1987.
12. Hertz, R. Hours of Service Violations Among Tractor-Trailer Drivers. *Accident Analysis and Prevention,* Vol. 23, No. 1, 1991, pp. 29–36.
13. Kaneko, T., 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.
14. Lin, T., P. P. Jovanis, and C. Yang. Modeling the Safety of Truck Driver Service Hours Using Time-Dependent Logistic Regression. In *Transportation Research Record 1407,* TRB, National Research Council, Washington, D.C., 1993, pp. 1–10.
15. Mackie, R., J. F. O'Hanlon, and M. A. McCanley. *A Study of Heat Noise and Vibration in Relation to Driver Performance and Physiological Status.* Report DOT HS-801 313. Human Factors Research, Inc., Goleta, Calif., 1974.
16. Mackie, R., 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.
17. Hertz, R. 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.
18. Sweedler, B. M., K. Quinlan, and M. Brenner. Fatigue and Drug Interaction in Fatal to the Driver Heavy Truck Crashes. *Proc., 34th Association for the Advancement of Automobile Medicine,* 1990, pp. 491–504.
19. Large Truck Survey Program. University of Michigan Transportation Research Institute, *Proc., of the National Truck Safety Symposium,* 1987.
20. Jovanis, P. P., 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.
21. Mathematics of Sleep. *Technology Review,* Feb.–March 1987.
22. Brown, C. C. On the Use of Indicator Variables for Studying the Time-Dependence of Parameters in a Response-Time Model. *Biometrics* 31, 1975, pp. 863–872.

23. Abbott, R. D. Logistic Regression in Survival Analysis. *American Journal of Epidemiology* Vol. 121, No. 3, 1985, pp. 465–471.
24. Efron, B. Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association*, Vol. 83, No. 402, 1988, pp. 414–425.
25. Hosmer, D. W., and S. Lemeshow. *Applied Logistic Regression.* John Wiley & Sons, Inc., N. Y., 1989.
26. Wu, M., and J. Ware. On the Use of Repeated Measurements in Regression Analysis with Dichotomous Response. *Biometrics*, Vol. 35, 1979, pp. 513–521.
27. Cupples, L. A., R. B. D'agostino, K. Anderson, and W. B. Kannel. Comparison of Baseline Repeated Measure Covariate Techniques in the Framingham Heart Study. *Statistics in Medicine*, Vol. 7, 1988, pp. 205–218.
28. Prentice, R. L., and R. Pyke. Logistic Disease Incidence Models and Case-Control Studies. *Biometrika* 66, 1979, pp. 403–411.
29. Yang, C., 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, 1992.

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