Title
Multiday Driving Patterns and Motor Carrier Accident Risk: A Disagregate Analysis

Authors
Kaneko, Tetsuya
Jovanis, Paul P.

Publication Date
1991-06-01

Multiday Driving Patterns and Motor Carrier Accident Risk: A Disaggregate Analysis

Tetsuya Kaneko
Paul P. Jovanis

June 1991
Working Paper, No. 59

The University of California Transportation Center

University of California
Berkeley, CA 94720

The University of California Transportation Center

The University of California Transportation Center (UCTC) is one of ten regional units mandated by Congress and established in Fall 1988 to support research, education, and training in surface transportation. The UC Center serves federal Region IX and is supported by matching grants from the U.S. Department of Transportation, the California State Department of Transportation (Caltrans), and the University.

Based on the Berkeley Campus, UCTC draws upon existing capabilities and resources of the Institutes of Transportation Studies at Berkeley, Davis, and Irvine; the Institute of Urban and Regional Development at Berkeley; the Graduate School of Architecture and Urban Planning at Los Angeles; and several academic departments at the Berkeley, Davis, Irvine, and Los Angeles campuses. Faculty and students on other University of California campuses may participate in Center activities. Researchers at other universities within the region also have opportunities to collaborate on selected studies. Currently faculty at California State University, Long Beach, and at Arizona State University, Tempe, are active participants.

UCTC's educational and research programs are focused on strategic planning for improving metropolitan accessibility, with emphasis on the special conditions in Region IX. Particular attention is directed to strategies for using transportation as an instrument of economic development, while also accommodating to the region's persistent expansion and while maintaining and enhancing the quality of life there.

The Center distributes reports on its research in working papers, monographs, and in reprints of published articles. For a list of publications in print, write to the address below.

University of California
Transportation Center

108 Naval Architecture Building
Berkeley, California 94720
Tel: 415/643-7378
FAX: 415/643-5456

Multiday Driving Patterns and Motor Carrier Accident Risk: A Disaggregate Analysis

Tetsuya Kaneko

Engineer
Sato Kogyo Company Ltd., Japan

Paul P. Jovanis

Department of Civil Engineering and Transportation Research Group
University of California at Davis

Working Paper, No. 59
June 1991

Previously published by the Transportation Research Group
University of California at Davis

The University of California Transportation Center
University of California at Berkeley

ABSTRACT

A method has been developed to estimate the relative accident risk posed by different patterns of driving over a multi-day period. The procedure explicitly considers whether a driver is on-duty or off-duty for each half hour of each day during the period of analysis.

From a data set of over 1000 drivers, 9 distinct driving patterns are identified. Membership in the patterns is determined exclusively by the pattern of duty hours for seven consecutive days; for some drivers an accident occurred on the eighth day while others had no accident, therefore each pattern can be associated with a relative accident risk. Additional statistical modeling allowed the consideration, in addition to driving pattern, of driver age, experience with the firm, hours off-duty prior to the last trip and hours driving on the last trip (either until the accident or successful completion of the trip).

The findings of the modeling are that driving patterns over the previous seven days significantly affect accident risk on the eighth day. In general, driving during afternoon and evening hours (e.g. noon to midnight) has the highest accident risk while driving during night and morning hours (e.g. midnight to noon) has lower risk. Consecutive hours driven also has a significant effect on accident risk: the first hour of driving and the ninth and tenth hour of driving have the highest risk. Hours 2 through 8 follow a flattened "u" shape. Driver age and hours off-duty immediately prior to a trip do not appear to affect accident risk significantly.

These findings quantitatively assess the relative accident risk of multi-day driving patterns using data from actual truck operations. Further research is recommended in the areas of refining model structures, adding explanatory variables (such as highway type) and testing more complex models.

I. INTRODUCTION

A. BACKGROUND

Interstate motor carriers are subject to limitations on the hours that their drivers may be on-duty and driving. They include a requirement that a driver be off-duty for a minimum of 8 hours after driving for 10 hours or being on-duty for 15 hours. There are also cumulative restrictions for on-duty time over several days: 70 hours on duty in 8 days for carriers operating 7 days a week and 60 hours in 7 days for those operating 5 days a week. These limitations, referred to as the hours of service regulations, were initiated in the 1930's. 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.

In a major book on fatigue, safety and the truck driver, MacDonald discusses the inconsistency and vagueness in how researchers have defined and used the concept of fatigue (MacDonald, 1984). For some it is subjective, dealing primarily with individuals' perceptions of how they feel. Others use physiological correlates or performance decrements to study fatigue (Haworth et. al., 1988).

There also appears to be confusion in some studies of the distinction between fatigue attributable to continuous driving and other time-related driving factors. Circadian rhythms are changes in body function that follow an approximate 24 hour period so there is a point of low rhythm which corresponds to a generally depressed level of arousal. In addition, sleep deprivation, which arises because of a combination of on-duty time and off-duty activities, may also influence arousal and, ultimately, accidents.

Fatigue is a sufficiently vague concept that it does not appear to be a useful focal point for this study. As an alternative, declines in performance as measured by accident risk are used as a measure of the quality of the driving task. Perhaps the most extensive studies of hours of service and accident risk were conducted in the 1970's as part of a series of studies sponsored by the National Highway Traffic Safety Administration (NHTSA), (Harris, et al., 1972; Harris, 1977; Mackie, et al., 1974; Mackie and Miller, 1978). These studies included analyses of retrospective accident data and field tests, with an instrumented cab, of a set of drivers asked to drive particular schedules. The effects of heat, noise, vibration and cargo loading activities were also assessed.

These studies consistently found that a higher proportion of accidents occurred in the last half of a trip. Separate analyses of single vehicle accidents and crashes for which the driver was reported as "dozing at the wheel" showed particularly strong increases in accident risk as a function of continuous hours driving. Circadian effects were significant for the "dozing" drivers as the accident risk was highest from 2-6 a.m. While there were some studies that included separate collection of exposure data, most of the analyses with accident data compared the actual number of accidents with those "expected" if there was no increased risk due to hours driven. This method assumes that accident involved drivers are representative of the general population of drivers. The studies also relied primarily on accident data from the then Bureau of Motor Carrier Safety (now the Office of Motor Carriers of Federal Highway Administration) although some data were provided directly from carriers. The studies do not explicitly consider the effect of total hours driven during preceding days nor the time of day when the driving occurred during those days.

There have been several recent studies that have explored aspects of accident risk and driving hours. The Insurance Institute for Highway Safety (IIHS) recently completed a study of drivers in sleeper berth operations (Hertz, 1989). While representing a rather unique type of over-the-road operations, Hertz found that regularity of schedule was an important predictor of road safety. In another study (Transportation Research and Marketing, 1985), a non-random set of accidents (primarily in the western U.S.) were selected for detailed follow-up. Interviews with firms and family members were used to reconstruct how the truck driver spent his time both on and off duty, in the day or so prior to the crash. The findings were that fatigue was a major contributing factor because of a combination of excessive (and illegal) hours of work and lack of rest during off-duty time. While the findings are of interest, the study suffers from methodological shortcomings: the criteria for selection of crashes would seem to be biased toward severe outcomes and the method used to determine the contribution of fatigue to accident occurrence seems subjective.

Studies have also been conducted in Europe. Hamelin, in an analysis of professional and non-professional drivers, found that the professional driver had lower accident rates than non-professionals, particularly during extended driving. He concluded that the professionals could better cope with the rigors of on-road performance (Hamelin, 1987). Fuller reached similar conclusions in his study of driving performance in Ireland (Fuller, 1981). No difference was found in the mean following headway of drivers, even after extended hours on-duty and driving.

Further research seeking to relate accident risk and motor carrier driving patterns can take any of several paths. One could seek to obtain detailed physiological and perceptual data from drivers undertaking truck driving tasks. This approach, best exemplified by Mackie and Miller (1978) is both costly and subject to criticism for its non-representativeness of actual driving conditions. An alternative is in-depth study of selected accidents (e.g. TRAM, 1985). This approach also can be questioned regarding generalizability. A third approach is to analyze accident data from actual truck operations, conduct some comparisons with non-accident events and seek to identify accident patterns that support or refute a relationship with driver hours regulations (this is much in the spirit of the research by Harris in 1972 and 1977).

The approach taken in this research is to follow the lead provided by Harris and his colleagues and seek to identify relationships between accident risk and driving service hours. In particular, this paper seeks to identify associations between continuous driving as well as driving patterns over a multiday period. The approach is predicated on the belief that, as it relates to safety policy, a primary concern is to focus on the effect of driving patterns on performance; i.e. a safely completed trip or an accident-producing trip. Rather than rely on information from accident reports or driver interviews that attempt to specifically attribute causality to factors such as fatigue, accident occurrence models are estimated in an attempt to test the potential effect on accident risk of actual driving patterns. By linking specific patterns to accident risk, it is hoped that high-risk as well as low-risk patterns will be identified. The linkage to real driving and on-duty time can then be related to existing and proposed hours of service regulations to assess their safety implications.

B. OBJECTIVES OF THE STUDY

The review of the literature suggests that there is a clear need to develop a method to analyze the effect on accident risk of different driver service hours. In particular, it is important to consider both the duration of on-duty and driving time just prior to the crash, as well as the times of day of driving over multiple days so that the cumulative effect of multiday driving can be assessed.

A second objective of the research is to test the method with data from trucking company operations. These should include data from accident reports as well as a comparable set of non-accident data so that relative accident risk can be assessed. Based upon previous research (Jovanis and Chang, 1989) it is highly desirable if the method includes accident and non-accident data at a disaggregate level.

II METHODOLOGY

A. WHAT IS A DRIVING PATTERN?

A driving pattern, for the purposes of this research, is a description of the status of the driver over several days. The driving patterns provide important information including: (i) hours on-duty and off-duty over multiple days; (ii) the time of day that the on-duty and off-duty hours occurred; and, (iii) trends of on-duty and off-duty time over several days. There are obviously a very large number of driving patterns that are 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 the effect of the pattern on risk can be assessed.

Statistical analysis of the driving patterns proceeds in two phases. First, there is a need for a method to extract similar driving patterns from a large pool. It is important that this determination of similarity be conducted in a way that is blind to accident occurrence. That is, the method should first group drivers with similar patterns; once similar patterns are identified, knowledge of the accident involvement of drivers with particular patterns can be used to assess accident risk.

Disaggregate non-accident trips present no problem in this regard. A non-accident driver is randomly selected from the population of drivers; a particular day is randomly selected and a trip (i.e. a ten hour or less driving period) is randomly selected within the day. Accidents are problematic because the occurrence of the accident interrupts the driving pattern, producing unknown biases. In order to avoid these biases the following approach is adopted. Driving patterns are described for the 7 days prior to the accident or comparable non-accident trip. This approach simplifies the statistical treatment of the data but results in the implicit assumption that the observed driving pattern over 7 days is carried into the eighth day. On the accident day or randomly selected non-accident day, driving is described by the number of hours driving since last consecutive 8 hours off-duty. As will be seen shortly, the patterns that result from this analysis are regular enough that this does not appear to be an unreasonable assumption. Driving patterns can be defined over any number of days. An eight day cycle is used because the carrier used in the empirical modeling operated seven days a week so the operative U.S.D.O.T. regulation limits drivers to 70 hours on-duty or driving in 8 days.

IDENTIFY DRIVING PATTERNS

- Use Cluster Analysis To Identify Similar Driving Patterns

- Select From Available Clusters Based Upon Ability to Discriminate Accident Risk While Meeting Sample Size Requirements

↓

USE ACCIDENT PATTERNS
AND OTHER PREDICTORS TO
ASSESS RELATIVE ACCIDENT RISK

- Test Range of Predictors for Logit Model
- Identify Preferred Model Based Upon Goodness of Fit and Parameter Significance

Table 1: Relationship of Statistical Methods

B. STATISTICAL METHODS

An overview of the statistical methods used in this research is contained in Table 1. Cluster analysis is a method that classifies objects by creating homogeneous groups. The driving pattern of an individual driver for the seven days prior to the day of the accident (or non-accident) is considered as the object; each driver is assigned to a cluster based upon the similarity of the driving pattern with other drivers in the cluster. The K cluster means option of BMDP is used to define the clusters. The default measure is used as the distance metric. Cluster analysis does not yield a single optimum set of clusters for a data set. The user has the option to select the number of clusters desired and the clustering algorithm assigns each observation to its most statistically similar cluster. A range of cluster numbers can be used, but some criterion is needed for selecting the clusters to be carried to the next step of the analysis.

The procedure used in this research tested a range of clusters from 5 to 9; the maximum number of clusters was determined by a rule of thumb that approximately 100 observations be contained in each cluster. The recommended clusters were determined by a Chi-squared test that sought to determine if the relative occurrence of accidents was independent of the number of clusters (n). A 2 by n contingency table was constructed and the clusters that best differentiated accident risk were chosen for further modeling. This is not an essential step as all clusters could have been tested in the second phase relative risk analysis. The screening seemed a reasonable method for reducing the number of models to be estimated.

There are two possible outcomes of each truck trip: an accident occurs or does not occur. This problem may be viewed as a binary response problem, which is a function of several factors. A particular binary response model, logit analysis, is applied to model the accident risk relative to the number of non-accident trips. Since not all non-accident trip records are collected or coded, the model should not be used to predict the actual probability of accident occurrence. However, it can be used to examine which factors contribute to accident occurrence more than other factors, thus the use of the term <u>relative</u> accident risk.

Note, non-accident drivers may appear more than once in the data if so selected by chance. Accident involved drivers may also appear more than once although this is unlikely; the company employs several thousand drivers nationwide. While bias may occur because an accident driver can appear more than once, the results reflect the policies of the firm and, at the level of the firm, are unbiased.

C. Data Used to Identify Driving Patterns

The primary source of data for this study are logs which describe driver duty status for each 15 minutes for accident and non-accident involved drivers. If a 7-day interval is considered, the number of variables are 672 (4 time periods per hour x 24 hours x 7 days). Computer memory limitations dictate that the finest time resolution that can be used is 30 minutes, decreasing the number of variables to 336 (2 x 24 x 7). The methods used to transform the 15 minute data to 30 minute intervals are as follows:

- If both 15 minute intervals have the same working status, the new variable (30 minute interval) has the same working status.

- On-duty and driving, and on-duty and not driving are treated as one working status which is on-duty (this is consistent with DOT cumulative hours regulations).

- If one of two 15 minute intervals is off-duty and another is on-duty, the new variable is treated as off-duty.

The last transformation will cause an underestimate of the hours on-duty, used as input to the cluster analysis. The output of the cluster analysis is distinct enough that little bias appears to have occurred. The transformation results in no error in estimating relative accident risk because it is used only to define driving patterns, not to measure driving, and off-duty times in the logit model.

The logit analysis is conducted using variables which include the age of the driver, the experience of the driver, the consecutive hours of driving on the trip in question, and the consecutive hours off-duty prior to the last trip. The total number of observations used in this study is 1419, however, since there are some missing values in age and experience, the total number of observation used for modeling is 1066 cases. The number of accidents is 382 and the number of exposure trips is 684.

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 USDOT service hour regulations, the vast majority of drivers in the study can be considered as operating within legal duty hour limits. The empirical results are intended as a test of the proposed methodology and as a contribution to the admittedly scant research base on accident risk and driving patterns.

III RESULTS OF DATA ANALYSIS

A. Analysis of Individual Driving Patterns

After experimenting with 5 - 9 clusters to describe driving patterns, the cluster analysis with 9 homogeneous driving patterns was used for further modeling. This decision was based on two factors. First, the 9 cluster results most clearly identified groups with dissimilar accident risks. The null hypothesis of independence was rejected of the 2 x 9 contingency table at the .05 significance level (see Table 2). The second factor is that the 9 cluster patterns appeared to be the most clearly discrete. Cluster analysis allocates observations to clusters based upon their statistical distance from cluster centroids. As observations are added to the clusters, the centroids shift, possibly resulting in previously assigned observations being misclassified. In all previous clusters (i.e. 5-8) there was a need to conduct 2 consecutive cluster analyses with the output of the first cluster being used as input to the second. In no case was more than 2 runs necessary. The 9 cluster case was the only one that did not result in any re-assignment of observations. Computer memory requirements and sample size restrictions prohibited testing of 10 or more clusters.

Figure 1 shows the overall average driving plan while Figure 2 - 10 represent individual clusters. The horizontal scale represents the elapsed time for each of the seven 24-hour periods. The time scale starts at midnight (point 0) and runs to 24 hours for the first day; 24-48 is the second driving day and so on until 144-168 representing the seventh driving day, just prior to the accident day. The vertical scale represents the percentage of drivers within the pattern that were driving or on-duty at that time. For example, in Figure 2, about 30% of drivers in pattern 1 are on-duty at midnight at the end of the first day (hour 24). The percentage of drivers on duty then drops to about 10% at 6:00 am on the second day (hour 30).

TABLE 2

Contingency Table of Cluster Number and Outcome Status

Cluster	Status		Total
	Non-Accident	Accident	
1	58	42	100
2	79	35	114
3	68	45	113
4	97	46	143
5	85	27	112
6	60	38	98
7	75	44	119
8	77	61	138
9	85	44	129
Total	684	382	1066

Statistics

Statistic	Value	d.f.	p-value
Pearson $\chi 2$	15.9	8	0.043

What is most startling about the figures is the difference in interpretation that is possible when comparing the aggregate pattern (Figure 1) to the individual clusters. Figure 1 merely reflects for this firm what has been commonly reported elsewhere for the industry as a whole. Truck drivers are on-duty throughout the day, for all seven days but there is a slight increase in the percentage of drivers on-duty in the evening and early morning hours (centered around midnight from about 6 pm until 8 am). Overall, the change in drivers on duty is from slightly more than 30% at midnight of the seventh day to a low of about 22% around noon of days 3,4,5,6.

Individual driving patterns are quire clearly identified using the clustering technique. In addition to a summary of the on-duty trends for each cluster a relative accident rate is reported. The relative accident risk associated with each cluster, n, is calculated as:

$$\text{Relative Accident Risk n} = \frac{a_n}{a_n + e_n}$$

where: a_n = number of trips resulting in an accident in cluster n
e_n = number of trips resulting in no accident in cluster n
n = the cluster number.

The interpretation of each individual driving pattern is as follows:

- Pattern 1

The most frequent driving periods in this pattern occur from early afternoon (about 3 p.m.) until about midnight but frequently extending until 3 - 4 a.m. Off-duty hours are thus most frequent from 4 a.m. until noon. Driving is somewhat irregular for the first 4 days of the pattern but is quite regular over the last 3 days; for example, over 80% of the drivers are on duty at 10 p.m. of the sixth day. This driving pattern is associated with a somewhat high level of accident risk, a relative accident risk of 0.420.

- Pattern 2

The most frequent driving periods in this pattern occur from early morning (about 2 a.m.) until slightly before noon. Off-duty times occur from early afternoon until near midnight. Driving is somewhat irregular during the first 4 days of this pattern but highly regular for the last three with steep peaks; for example, nearly 75% of the drivers are on duty at 11 a.m. on the sixth day. This driving pattern is associated with a somewhat high level of accident risk, a relative accident risk of 0.307.

- Pattern 3

The most common on-duty hours in this pattern are in the morning, beginning after midnight and extending until nearly noon. The most common off-duty time is noon to midnight. Driving becomes very infrequent during the last 2 days of the pattern but is highly regular during the first 5 days; for example, on the fourth day nearly 80% of the drivers are on-duty at about 6 a.m. This pattern is associated with moderate accident risk, a relative accident risk of 0.398.

- Pattern 4

The most frequent on-duty hours in this pattern are from morning, about 6 a.m., through the afternoon, until about 6 p.m. Hours are very regular for the first 3 days but somewhat less so during day 4 and even less so during 5. Driving is rather unlikely during days 6 and 7. Off-duty hours typically occur from evening (about 6 p.m.) through early morning (about 6 a.m.). Nearly 80% of the drivers in this group are on-duty at noon on the first and second days. This pattern is associated with rather low level of accident risk, a relative accident risk of 0.322.

- Pattern 5

The most frequent on-duty time for this group of drivers occurs from early evening, around 8 p.m., through early morning, about 6 a.m. Off-duty times are typically late morning through early afternoon. The pattern is highly regular during the first 2 days (more than 80% of the drivers on-duty at the beginning of the second day) and somewhat less so during days 3, 6 and 7. The least frequent on-duty days are 4 and 5. This pattern is associated with the lowest level of accident risk, a relative accident risk of 0.241.

- Pattern 6

This pattern contains drivers that are very infrequently scheduled, particularly during the first 6 days. On the seventh day, only 30% of the drivers in this pattern are on-duty from mid-night until about 6 a.m. This pattern is associated with moderate accident risk, a relative accident risk of 0.388.

- Pattern 7

The most frequent on-duty times for drivers in this group are from about noon until about 10 p.m. The most likely off-duty time is from midnight until about 10 a.m. The pattern is quite regular on the last 3 days of the 7 day period with nearly 80% of the drivers on-duty during day 6 and somewhat less so during days 5 and 7. The first 4 days of the pattern demonstrate somewhat more variability but there is a pronounced peak period as typically 40% or more of the drivers are on-duty during the peak time. This pattern has a moderate relative accident risk of 0.340.

- Pattern 8

The most frequent driving times start at about 10 p.m. and continue through about 10 a.m. The most frequent off-duty times are 10 a.m. through about 10 p.m. The pattern is highly regular during the last 4 days with a peak of 70 % of the drivers on-duty on days 5, 6 and 7. The first 3 days exhibit much higher variability. This pattern has the highest accident risk, a relative accident risk of 0.442, in our data set.

- Pattern 9

The most frequent on-duty time for these drivers is throughout the afternoon and evening from about 6 p.m. until just before midnight. The most likely off-duty time is late morning and early afternoon. The most frequent on-duty days are 1 through 5 but there is much less-pronounced peaking within this pattern. This pattern is associated with rather low accident risk, a relative accident risk of 0.341.

By inspecting the clusters, several common trends emerge. Patterns 1, 2, 7 and 8 all contain relatively infrequent irregular driving during the first 3 - 4 days but highly regular driving thereafter. This is derived from, for example, the observation that 40% or less of the drivers in pattern 1 are on-duty or driving from about noon to midnight on days 1 - 4 but this percentage rises to 70% on day 5 and 7 and 80% on day 6. Conversely, patterns 3, 4 and 9 have regular driving during days 1 - 4 and more irregular driving thereafter.

Several sets of patterns have similar peak hours of driving within the day, but differ principally in which days during the 7 day period exhibit irregular duty hours. For example, both pattern 1 and 9 contain peak driving from early afternoon (e.g. 3 p.m.) until very early morning (e.g. 3 a.m.). The major difference is that pattern 1 has irregular duty hours on the first 4 days while pattern 9 exhibits irregular duty hours on days 5 - 7. This "phase shift" is also apparent in comparisons of patterns 2 and 3, 4 and 7, 5 and 8. Accident risks for each of these phase shift patterns are specifically compared in one part of the next section.

B. Logit Model

The logit model is calibrated in order to estimate the relative accident risk, which is the chance of an accident out of the total number of trips (which are the sum of the accident trips and the exposure trips).

The logit formulation of this case is expressed as follows;

$$P_i(\text{Accident}) = \frac{e^{\alpha+\beta_1 X_{i1}+\cdots+\beta_n X_{in}}}{1 + e^{\alpha+\beta_1 X_{i1}+\cdots+\beta_n X_{in}}} \tag{1}$$

$$P_i(\text{Exposure}) = \frac{1}{1 + e^{\alpha+\beta_1 X_{i1}+\cdots+\beta_n X_{in}}} \tag{2}$$

where i indicates each individual driver,
α is specific constant for accident occurrence,
$\beta_1 \cdots \beta_n$ are the parameters for each variable,
and $X_{i1} \cdots X_{in}$ are the variable for each individual i.

(Note that all the variables are included in the linear function in the numerator's exponent for accident occurrence and no variables for exposure.)

The advantage of the formulation is the convenient exponential form. If the linear function (combination of variables and associated weights) is viewed as a function expressing the accident risk which each driver experiences, then the coefficient associated with each variable represents the contribution of each variable to accident risk. As this individual propensity increases, the probability of accident occurrence for that specific individual increases. Positive coefficients associated with each variable indicate positive contribution to a higher probability of accident occurrence.

The variables used in this study are age, experience, driving pattern, consecutive driving hours on the last trip, and consecutive off-duty hours prior to the last trip. In order to include possible non linearities in the explanatory variables, a series of dummy variables expressing the categories of each variable are defined and summarized in Table 3. For each explanatory variable a base value is defined for which a parameter is not estimated; for each set of dummy variables the number of parameters is one less than the number of categories. The age of drivers is divided into three categories; the first category is the base. The experience of drivers is divided into 4 categories, more than ten years experience is the base. The driving patterns were categorized into the 9 homogeneous groups from the cluster analysis; pattern 5 is used as a base. The variable definition for driving hour is strictly correct for accident outcomes; i.e. for each accident only one driving hour category was coded as 1, all others were zero. For non-accident trips, however, a 1 is coded if the trip lasted at least as long as the category. Thus a 7 hour non-accident trip would have a 1 in each of the first 3 driving hour categories and zeros thereafter. This is believed to be the only way that the logit model can be made to accurately assess the risk of a trip of a given duration. The consecutive driving hours is divided into 5 categories; less than one hour driving is the base. The consecutive off-duty hours prior to the last trip is divided into 4 categories; 8-9 hours off-duty is the base. Higher order effects have been excluded from the scope of this model, however, the importance of higher order effects is not undervalued and has been identified for further study in the continuation of this research. Note also that driving greater than 9 hours is estimated to have a coefficient indistinguishable from zero, so it appears in Table 3 but not in Table 4.

The boundaries of the explanatory variables are selected based upon a desire to test particular hypotheses and a review of values typically used in truck safety studies. For example, it is of interest to determine the relative risk of drivers in their first year of driving. Similarly, it is of interest to determine accident risk during the first hour of continuous driving. Remaining categories are selected to balance the number of observations in the remaining categories.

The best model is obtained by a stepwise procedure (Kaneko, 1990) and the estimation results of the best model are reported in Table 4. The coefficient estimates, the asymptotic standard errors, and the t-statistics for each variable included in the model are presented. Moreover, several test statistics are presented as summary statistics.$L(\mathbf{0})$ is the value of the log likelihood function when all the parameters are zero, $L(\mathbf{c})$ is the value of the log likelihood function when only the specific constant for accident occurrence is included, and $L(\beta)$ is the value of the log likelihood function at its estimated maximum. Notice that β indicates the vector of the maximum likelihood estimates of the parameters.

TABLE 3

Variables used in Model Formulation

Variable Name	Definition of Variable
Age	
Age ≤40	1 if age of driver is less than or equal to 40 years old
40<Age≤50	1 if age of driver is more than 40 and less than or equal to 50 years old
50<Age	1 if age of driver is more than 50 years old
Experience	
Experiences≤1	1 if experience of driver is less than or equal to 1 year with the firm
1<Experience≤5	1 if experience of driver is more than 1 and less than or equal to 5 years with the firm
5<Experiences≤10	1 if experience of driver is more than 5 and less than or equal to 10 years with the firm
10<Experience	1 if experience of driver is more than 10 years with the firm
Driving Pattern	
Pattern 1	1 if driver belongs to pattern 1
Pattern 2	1 if driver belongs to pattern 2
Pattern 3	1 if driver belongs to pattern 3
Pattern 4	1 if driver belongs to pattern 4
Pattern 5	1 if driver belongs to pattern 5
Pattern 6	1 if driver belongs to pattern 6
Pattern 7	1 if driver belongs to pattern 7
Pattern 8	1 if driver belongs to pattern 8
Pattern 9	1 if driver belongs to pattern 9
Driving Hour	
DriveHour≤1	1 if driving hour is less than or equal to 1 hour
1<DriveHour≤5	1 if driving hour is more than 1 and less than or equal to 5 hours
5<DriveHour≤7	1 if driving hour is more than 5 and less than or equal to 7 hours
7<DriveHour≤9	1 if driving hour is more than 7 and less than or equal to 9 hours
9<DriveHour	1 if driving hour is more than 9 hours
Off-Duty Hours	
8≤OffDutyHour≤9	1 if off-duty hour is more than 8 and less than or equal to 9 hours
9<OffDutyHour≤12	1 if off-duty hour is more than 9 and less than or equal to 12 hours
12<OffDutyHour≤24	1 if off-duty hour is more than 12 or less than or equal to 24 hours
24<OffDutyHour	1 if off-duty hour is more than 24 hours

TABLE 4

Estimation Results

Summary Statistics

Variable Name	Coefficient Estimate	Asymptotic Standard Error	t statistic
Constant (Accident)	1.82670	0.37587	4.85990
Experience≤1	0.66109	0.29214	2.26290
1<Experience≤5	1.10600	0.28898	3.82730
5<Experience≤10	0.49046	0.23611	2.07720
Pattern 1	1.98780	0.43779	4.54050
Pattern 2	1.11130	0.44430	2.50110
Pattern 3	1.01440	0.39847	2.54580
Pattern 4	1.63890	0.40082	4.08880
Pattern 6	1.09680	0.40998	2.67530
Pattern 7	1.68150	0.43137	3.89810
Pattern 8	1.01840	0.37533	2.71340
Pattern 9	1.30550	0.39712	3.28730
1<DriveHour≤5	-3.70000	0.25638	-14.43200
5<DriveHour≤7	-2.22510	0.23783	-9.35560
7<DriveHour≤9	-1.55510	0.29151	-5.33470

Summary Statistics

Number of Observations	1066
Number of Parameters	15
Number of Iterations	7
L(0)	-738.90
L(c)	-695.52
L(β)	-356.11
-2[L(0)-L(β)]	765.58
-2[L(c)-L(β)]	678.83
ρ2	0.49
$\bar{\rho}^2$	0.47

According to Ben-Akiva and Lerman [1987], the following test statistics should be computed.

- The log test statistic, '$-2[L(\mathbf{0})-L(\beta)]$', which is used to test the null hypothesis that all the parameters in the model are zero, is asymptotically distributed as chi-square distribution with K degrees of freedom (where K is the number of parameters). For the model in Table 4, the number of parameters is 15, and the value of this statistic is 765.58. Since the theoretical chi-square value with 15 degrees of freedom is 24.996 at $\alpha = 5\%$ significance level, the null hypothesis can be rejected. Thus some variables should be included in the model.

- Another log test statistic, '$-2[L(\mathbf{c})-L(\beta)]$', which is used to test the null hypothesis that all the parameters other than the constant for accident occurrence are zero, is asymptotically distributed as chi-square distribution with (K-1) degrees of freedom. For the model in Table 4, the number of parameters (K) is 15, and the value of this statistic is 678.83. Since the theoretical chi-square value with 14 degrees of freedom is 23.685 at $\alpha = 5\%$ significance level, the null hypothesis can be rejected. Thus not only the constant but also other variables should be included in the model. Note that $L(\mathbf{c})$ is the likelihood obtained when the probability model replicates the overall average relative accident rate.

- ρ^2 which is defined as '$1-[L(\beta) / L(\mathbf{c})]$' is a measure similar to R^2 which is used in regression as a goodness-of-fit index. The ρ^2 value equals 0.49, which means that the goodness-of-fit of this model is almost 50%; however, this ρ^2 is an informal measurement. Hence an additional analysis of residuals is appropriate to verify this apparent goodness of fit of the model.

- $\bar{\rho}^2$ which is defined as '$1-\{[L(\beta)-K] / L(\mathbf{c})\}$' is similar to ρ^2, but is corrected for the number of estimated parameters. Here $\bar{\rho}^2 = 0.47$, and it is slightly smaller than ρ^2.

The model is interpreted by inspecting the coefficient estimates for each variable. Before examining the coefficient estimates of the best model, the variables which are excluded during the stepwise procedure are considered. Comparing the variables in Table 3 and the variables in the best model, all variables describing driver's age and consecutive off-duty hours prior to the last trip are excluded during the stepwise procedure. It is likely that the effect of the age of the drivers is not easily captured in the model as it is correlated with experience. It may be that age will appear only in interaction with other variables. The effect of consecutive off-duty hours prior to the last trip may be explained by the patterns since many of these off-duty hours may be part of the seventh day of the patterns.

The coefficient estimates in the best model are plotted in Figures 11 to 13: Figure 11 for experience; Figure 12 for driving patterns; Figure 13 for consecutive driving hours on the last trip. Model parameters are compared on the basis of their statistical significance using a T test. The null hypothesis is that the underlying true parameters are equal; i.e. $\beta_1 = \beta_2$. A test for a significant difference between β_1 and β_2 is based upon the following test statistic that is assymptotically T distributed:

$$T = \frac{\hat{B}_1 - \hat{B}_2}{\hat{\sigma}} \tag{3}$$

where $$\hat{\sigma}^2 = \hat{\sigma}^2_{\beta_1} + \hat{\sigma}^2_{\beta_2} - 2\left[\mathrm{Cov}\,\hat{\beta}_1, \hat{\beta}_2\right] \tag{4}$$

All $\hat{\beta}_1$'s and $\hat{\sigma}$'s are shown in Table 4. Covariance are obtained from the output of the logit estimation. The interpretation for each set of coefficients follows;

- Experience -

The coefficient estimate of '10<Experience' is set as zero, which is intended to illustrate that drivers with greater than ten years experience have the lowest relative accident risk. Model estimates confirm this expectation as all remaining categories of the experience variable are positive and included in the model; the conclusion is that drivers with more than 10 years experience are significantly safer than those at other experience levels. The drivers with 1-5 years experience have significantly higher relative risk than those with 5-10 years experience, but are indistinguishable in relative risk from the group with less than 1 year experience. This is somewhat surprising as the least experienced group was expected to have the highest relative risk. Two factors may intervene: experience is recorded as years with the firm (not driving trucks). Since this firm frequently hires drivers from other companies, the least experienced with the firm are not totally new to motor carriage. Second, it is possible that we are underestimating the risk for the drivers with less than 1 year experience as these may be more likely to be missing from our age/experience lists.

- Driving Pattern -

The coefficient estimate of pattern 5 is set as zero as it was expected to have the lowest accident risk based upon the cluster analysis. This expectation is confirmed as all patterns other than 5 are entered into the model (and are thus statistically significant) with positive parameters. Pattern 1 has the highest coefficient estimate, and the order of decreasing coefficient estimates for the other patterns is: pattern 7, pattern 4, pattern 9, pattern 2, pattern 6, pattern 8, and pattern 3. Generally, these patterns are categorized into two sets: One set consists of pattern 1, 4, 7, and 9, which present relatively high coefficient estimates and thus relatively high accident risk. Another set consists of pattern 2, 3, 6, and 8, whose coefficient estimates are statistically indistinguishable from 1 indicating that these patterns have an accident risk 100% higher than the baseline pattern. Despite the apparent differences in parameter values, statistical tests of the null hypothesis of equality of 2 parameters fails to reject the null except for marginal rejection of Pattern 1 compared to all patterns except 4 and 7 (i.e. H_0 is rejected at $\alpha = .10$ but accepted at $\alpha = .05$). All other pairs of patterns fail to reject the null hypothesis of parameter equality.

The first set (patterns 1, 4, 7, and 9) contain primarily afternoon and evening driving patterns. Driving most often occurs between noon and midnight for these drivers. Three patterns in the second set (patterns 2, 3, and 8) contain most frequent duty hours between midnight and noon. Pattern 6 is a particular pattern in which the majority of drivers are off-duty for nearly the entire seven days.

In general, afternoon and evening driving patterns have higher accident rate than night time and early morning driving patterns. This result may be at least partially due to lower traffic levels on the highway during night and early morning.

Additional insight is obtained by comparing the accident risk of the pairs of patterns that appear similar except for 3 - 4 day phase shift. Recall that these phase shift pairs are 1 and 9, 2 and 3, 4 and 7, and 5 and 8. Examination of the coefficients in Figure 12 reveal that patterns which contain significant on-duty time during days 5 - 7 (these are patterns 1, 2, 7 and 8) have a higher accident risk than the comparable paired patterns (i.e. 9, 3, 4 and 5) which have off-duty time during days 5 - 7. There thus does appear to be evidence of increased accident risk with cumulative driving that occurs over several driving days even for similar times of day. It is clear, however, that this effect is not consistent across all pairs: the pairs with patterns 2 - 3 and patterns 4 - 7 show very small accident risk differences while pattern 1 - 9 and 5 - 8 are quite large. Statistical tests of the parameters of the phase shift pairs, however, reveal marginally significant differences only between coefficients for patterns 1 and 9. While statistical tests do not fully support the qualitative conclusions apparent from comparison of parameters, the exploratory nature of the research must be considered. More empirical tests are needed to support or refute other hypotheses about multiday driving risk.

- Driving Hours -

The coefficient estimate for the first hour of driving, 'DriveHour≤1', was set as zero in the model to establish a baseline accident risk. The pattern of coefficients, illustrated in Figure 13, show a "u" shape. Accident risk is the highest in the first hour of driving, drops to the lowest value in hours 1 - 5 and gradually increases until the risk beyond 9 hours equals the risk in the first hour. At least part of the reason for the beginning hour having a higher accident rate is that the drivers must leave the terminal which is usually located in a relatively higher risk local road before they can access the lower risk interstate highway used for the majority of their travel. Drivers may also need a "warm up" to familiarize themselves with the large combination vehicle.On the other hand, the reason for higher accident rate beyond 9 hours of driving period may be a combination of increased driving time and local access. What is of interest here is that this u-shaped driving risk is retained even when controlling for experience and detailed driving patterns over the previous 7 days. This is the strongest evidence to date that the driving risk does gradually increase as driving time increases, but that the increased risk is due to a combination of local access and extended driving time.

An examination of the magnitude of the logit model coefficients in Table 3 indicates that consecutive driving hours exert an extremely strong influence on accident risk. The coefficient values of -3.7 and -2.2 for driving 1-5 and 5-7 hours consecutively are the highest magnitude coefficients by far. Each of the driving hours parameters are significantly different from each other.

IV. SUMMARY

Driver hours of service regulations limit the number of consecutive hours driving and on-duty as well as cumulative on-duty time over several days. A statistical procedure, using cluster analysis and discrete logit models, has been developed and tested on a data set reflecting actual motor carrier accidents and travel exposure. The data set represents the operating experience of one firm and should not be generalized. The findings do confirm the feasibility of the method, particularly for analysis of LTL operations. In addition, the findings are useful because they do appear to be very consistent with psychological theories.

The analysis of carrier-supplied data revealed that the pattern of driving over 7 days that was associated with the highest accident risk was daytime and early evening driving. Night and early morning driving over 7 days posed the least risk. Drivers did appear to be directly affected by cumulative driving over several days: for each of 4 pairs of similar driving patterns the ones that contained driving on the sixth or seventh day had a consistently higher accident risk on the eighth day than those with day 6 and 7 off-duty. Statistical tests, however, indicated that all patterns have statistically indistinguishable risks except for patterns 1 and 5. Pattern 1 had a marginally higher risk than all but two of the patterns. Pattern 1 consisted of drivers who typically drove from 3pm to 1am and drove during the 2-3 days prior to the day of interest. The safest pattern (5) consisted of driving from 10pm to 8am with driving 1-2 days prior to the day of interest but off duty time 3-4 days before. While other driving patterns yielded clearly interpretable relative risks, statistical tests failed to distinguish between the risks of any of the remaining patterns.

Consecutive hours of driving were the most significant predictor of accident risk. The first and tenth hour of driving had the highest (and equal) level of relative accident risk. The risk dropped precipitously in the second hour then rose through 2-5, 5-7 and 7-9 hours. Driver experience was significantly associated with relative accident risk but driver age and hours off-duty were not.

The findings reveal that it is possible to quantitatively account for both the number of hours on duty and driving over a seven day period as well as the time of day when the driving occurred. Consecutive hours driving are also strongly associated with accident risk but, unlike some previous studies, the first and tenth hour have the highest relative risks and the risks are equal.

Numerous additional analyses are possible with the existing data set or with enhancements made to the existing data. There is a need explore additional driving patterns and their effect on accident risk. While the 9 clusters in this study yielded very interpretable results, additional insights may be gained by trying to develop a larger number of clusters that are more precise in their driving patterns. This analysis requires additional data, beyond the 1066 cases used in this study. These studies need to be pursued as additional modeling may result in statistical significance for some of the parameters that were indistinguishable in this data set.

There is a need to include a broader range of variables that may help to explain accident occurrence. Some of these variables pertain to the routes used by the drivers: the road design, traffic level and terrain, for example. Individual driver socio-demographic characteristics such as marital status and family structure may also help explain accident risk.

There is a need for additional modeling. At present, all data are included in one model (the logit model) that explains accident risk. It may be that separate models should be constructed for each of the patterns. These systems of models may give an even clearer indication of accident risk. Another method of exploring the utility of additional models is to consider interaction terms in a logit or survival theory structure (Jovanis and Chang, 1989).

There is a need to more directly link the driving patterns on days 1-7 with driving on the eighth day. As in the pattern analysis, the coding of data on the eighth day should recognize the pattern of duty hours for each 1/2 hour throughout the day.

It is hoped that the use of cluster analysis in identifying multiday driving patterns will encourage similar studies with this methodology. Disaggregate analyses are becoming much more common in the truck safety literature (e.g. Chang and Jovanis, 1990; Leigh-Gosselin, et. al., 1990) and offer the prospect of more accurate identification of relative accident risk as well as the absolute probability of accident occurrence (Chang, 1987).

ACKNOWLEDGEMENTS

The authors wish to express their appreciation for financial support contributed by Yellow Freight Systems, Inc. and the University of California Transportation Center. The authors also wish to thank 2 anonymous referees for thorough reviews of the original text. Of course, the authors assume all responsibility for the content, ideas and recommendations of the paper and any errors it may contain.

REFERENCES

Ben Akiva, M. and S. Lerman. Discrete choice analysis: theory and application to travel demand, Cambridge MA., MIT Press, 1987.

Chang, H. A disaggregate survival model of motor carrier highway accident occurrence, Ph.D., Dissertation, Department of Civil Engineering, Northwestern University, Evanston, IL, 1987.

Chang, H. and P.P. Jovanis. Formulating accident occurrence as a survival process, Accident Analysis and Prevention, 22:407-419; 1990.

Fuller, R. Determinants of time headway adopted by truck drivers, Ergonomics, 24: 463-474; 1981.

Hamelin, P. Truck driver involvement in traffic accidents as related to their shiftworks and professional features, Symposium on Role of Heavy Freight Vehicles in Traffic Accidents, Vol. 2, Roads and Transportation Authority of Canada, Ottawa, Canada, April 1987.

Harris, W. Fatigue, circadian rhythm, and truck accidents, Vigilance: Theory, operational performance and physiological correlates, Plenum Press, New York, 1977, p.133-146.

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, Final Report, BMCS RD 71-2, Washington, D.C., U.S. Department of Transportation, Federal Highway Administration, Bureau of Motor Carrier Safety, November, 1972.

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, 124p.

Hertz, R. Tractor-trailer driver fatality: the role of nonconsecutive rest in a sleeper berth, Accident Analysis and Prevention, 20:431-439; 1988.

Jovanis, P.P. and H. Chang. Disaggregate model of highway accident occurrence using survival theory, Accident Analysis and Prevention, 21:445-458; 1989.

Kaneko, T. Multivariate analysis of motor carrier accident risk, M.S. Thesis, Department of Civil Engineering, University of California, Davis, California, 1990.

Lee-Gosselin, M.E.H., A.J. Richardson, and G. Taylor. Truck accident involvement with and without front axle brakes: an application for case-control methodology, Presentation, Transportation Research Board 69th Annual Meeting, Washington D.C., January 1990.

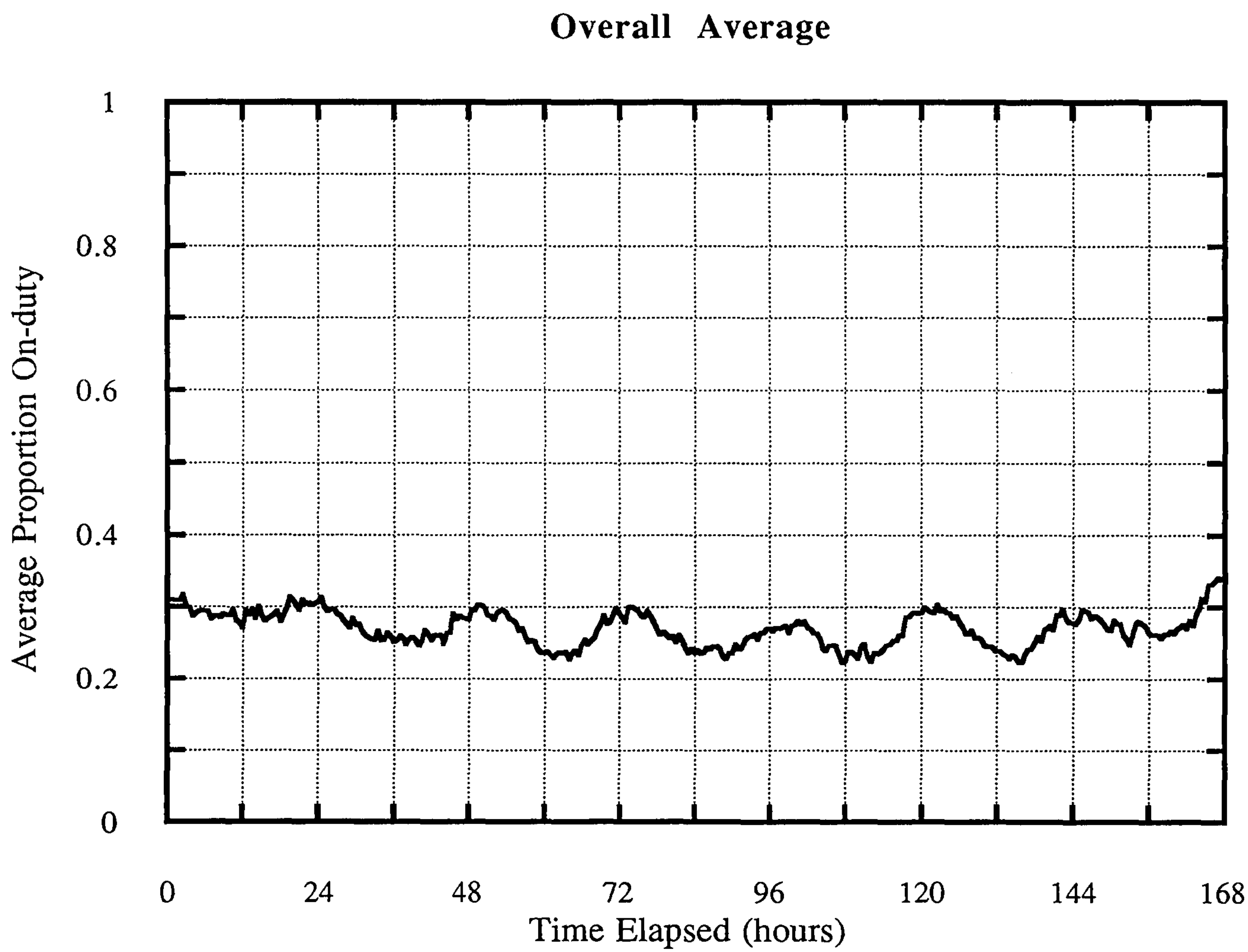
MacDonald, N. Fatigue safety and the truck driver, London, Taylor and Francis, 1984.

Mackie, R. and J. Miller. Effects of hours of service, regularity of schedules and cargo loading on truck and bus driver fatigue., Report No. DOT HS-803 799, Human Factors Research Incorporated Santa Barbara Research Park, Goleta, California, 1978.

Mackie, R. et al. 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.

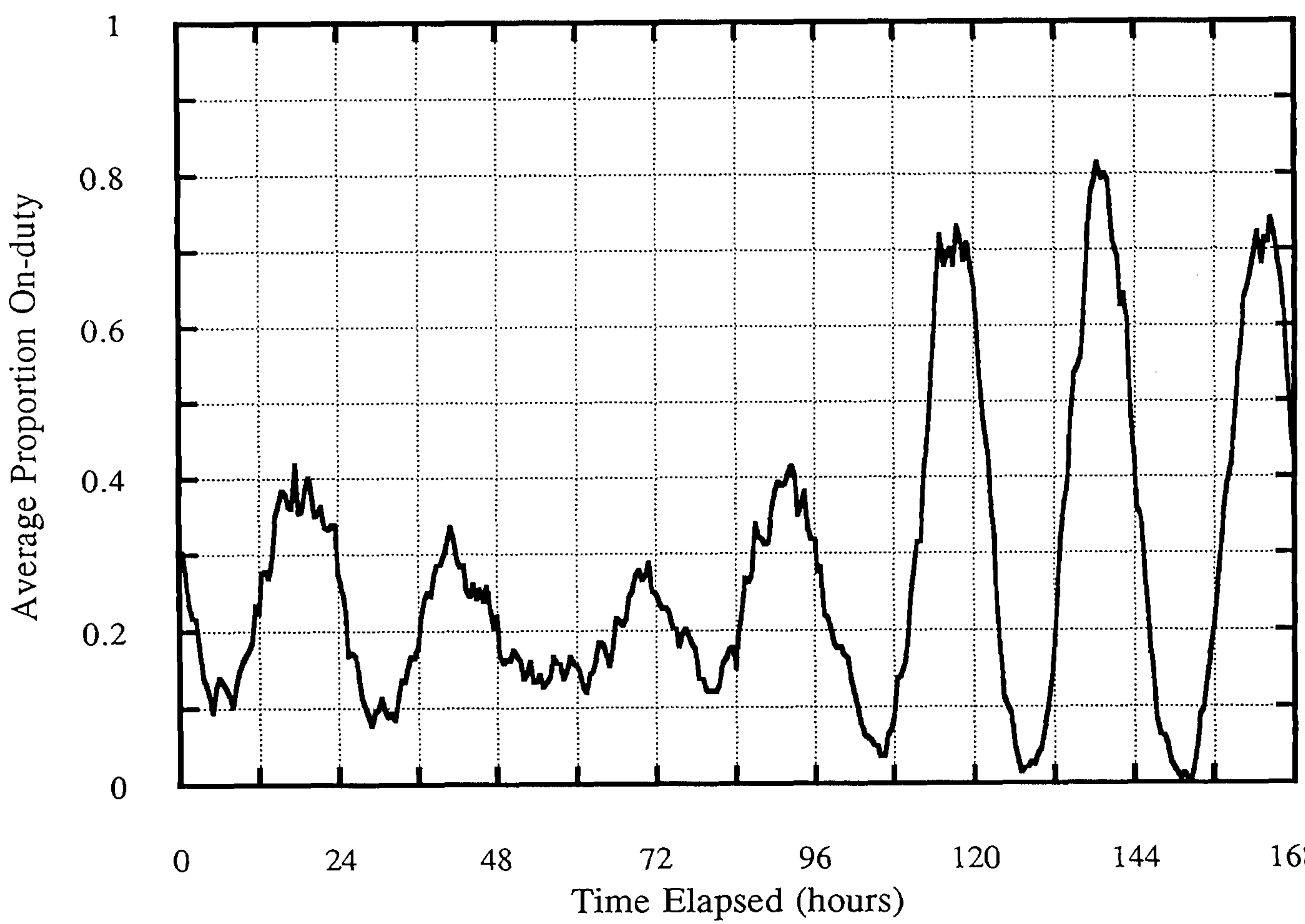
Transportation Research and Marketing. A report on the determination and evaluation of the role of fatigue in heavy truck accidents, AAA Foundation for Traffic Safety, Falls Church, VA, 1985, 76 p.

LIST OF FIGURE TITLES

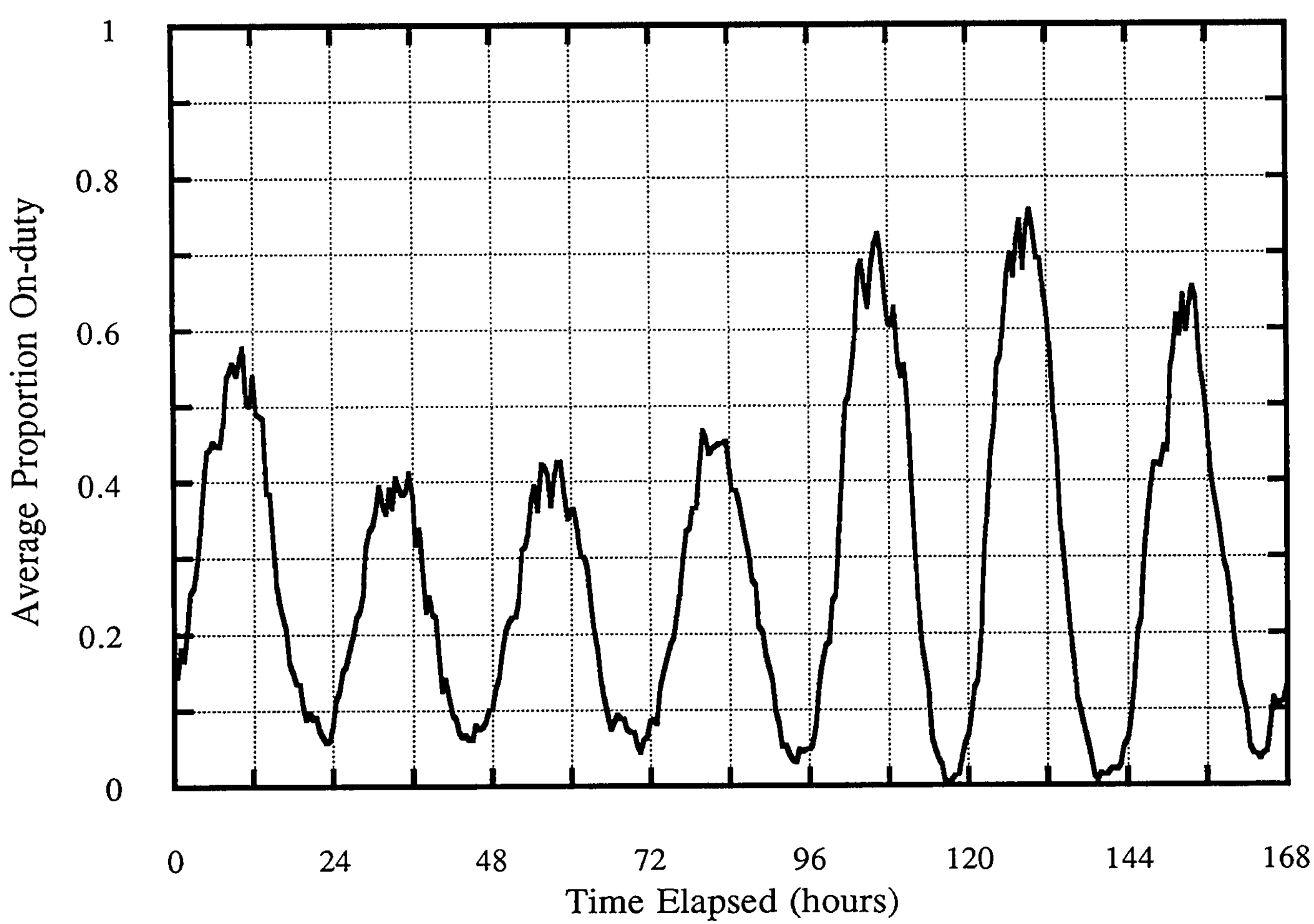


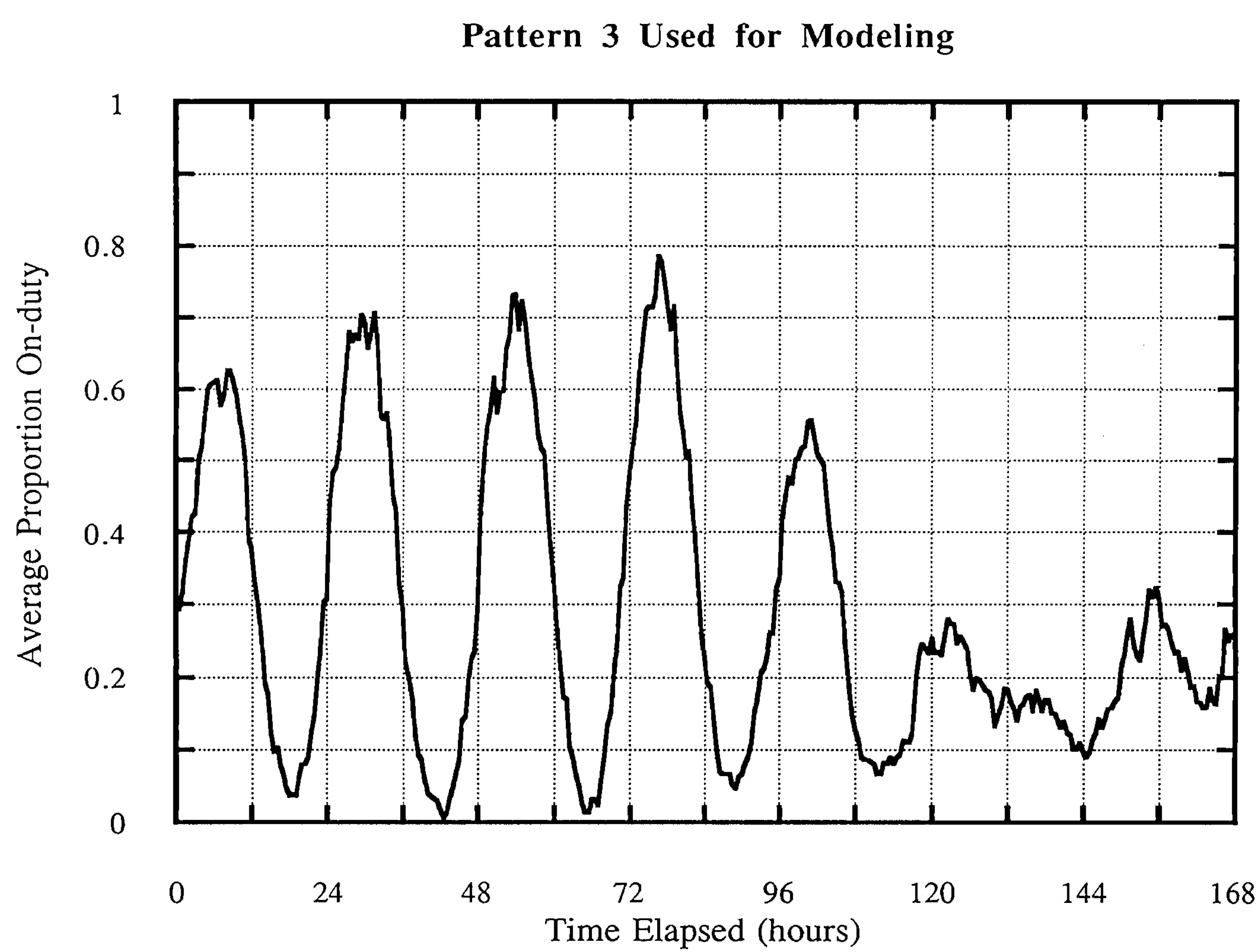
Overall Average
Average Proportion On-duty
1
0.8
0.6
0.4
0.2
0
0
24
48
72
96
120
144
168
Time Elapsed (hours)

Pattern 1 Used for Modeling

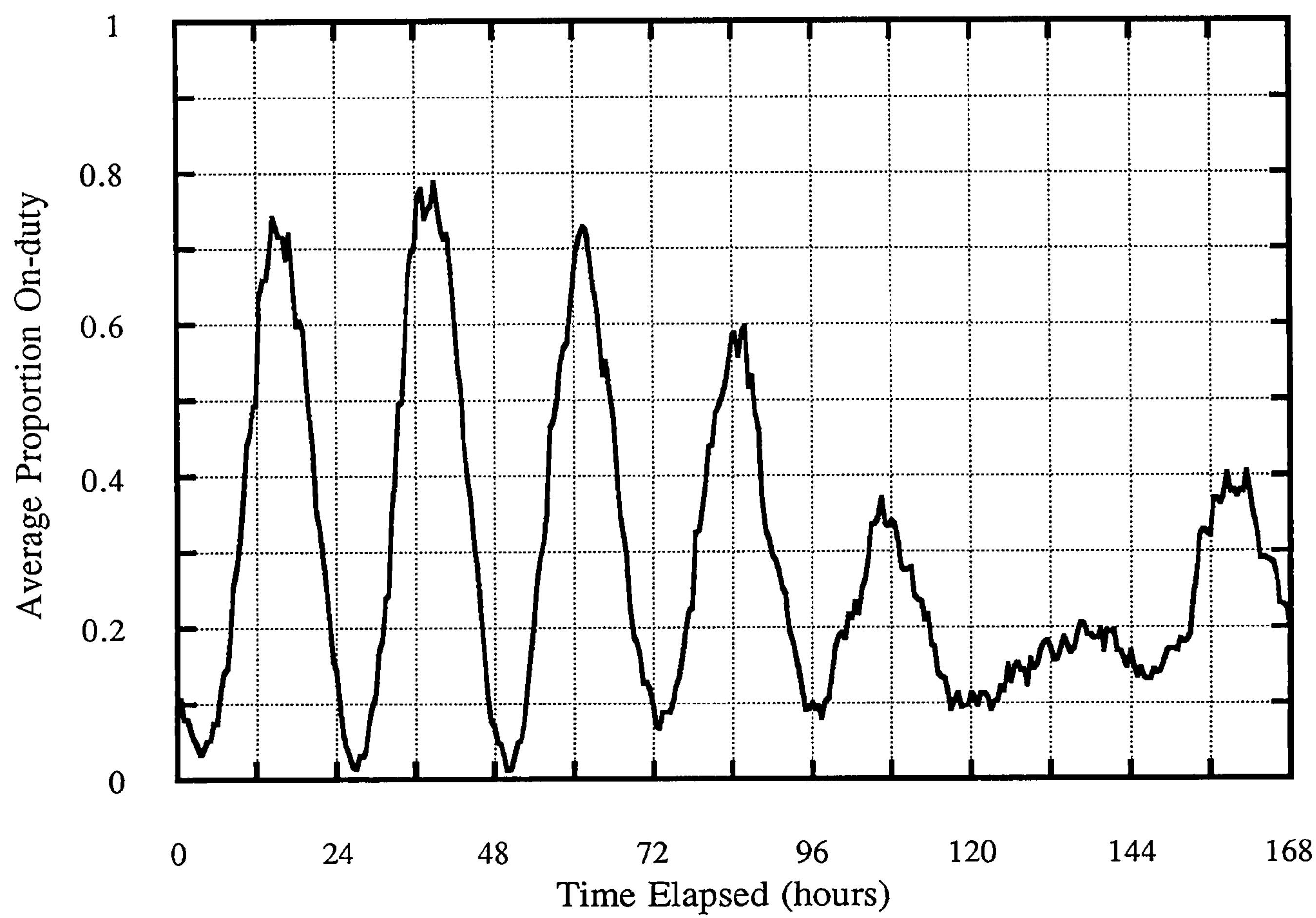


Pattern 2 Used for Modeling

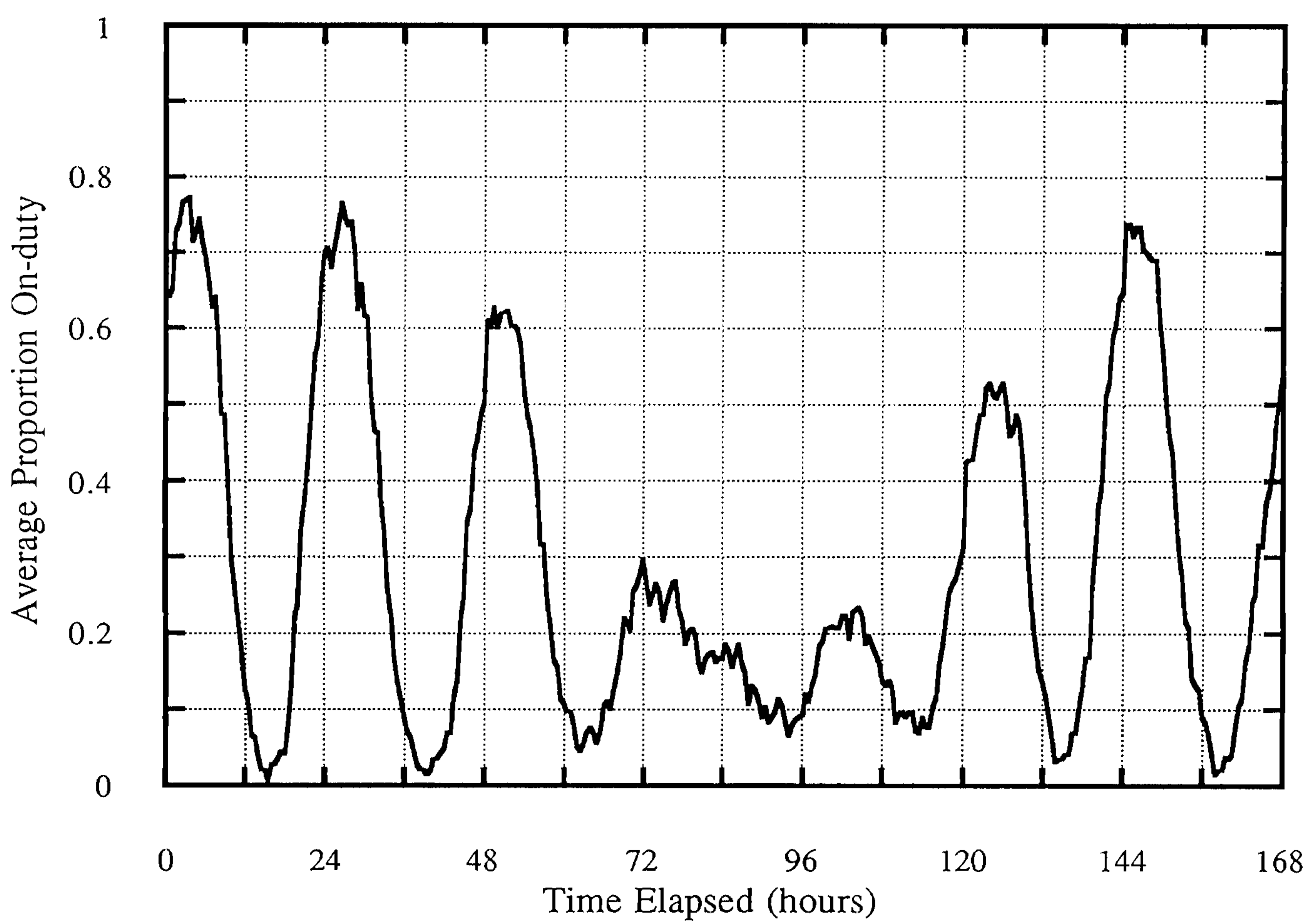



Pattern 3 Used for Modeling
Average Proportion On-duty
0
0.2
0.4
0.6
0.8
1
0
24
48
72
96
120
144
168
Time Elapsed (hours)

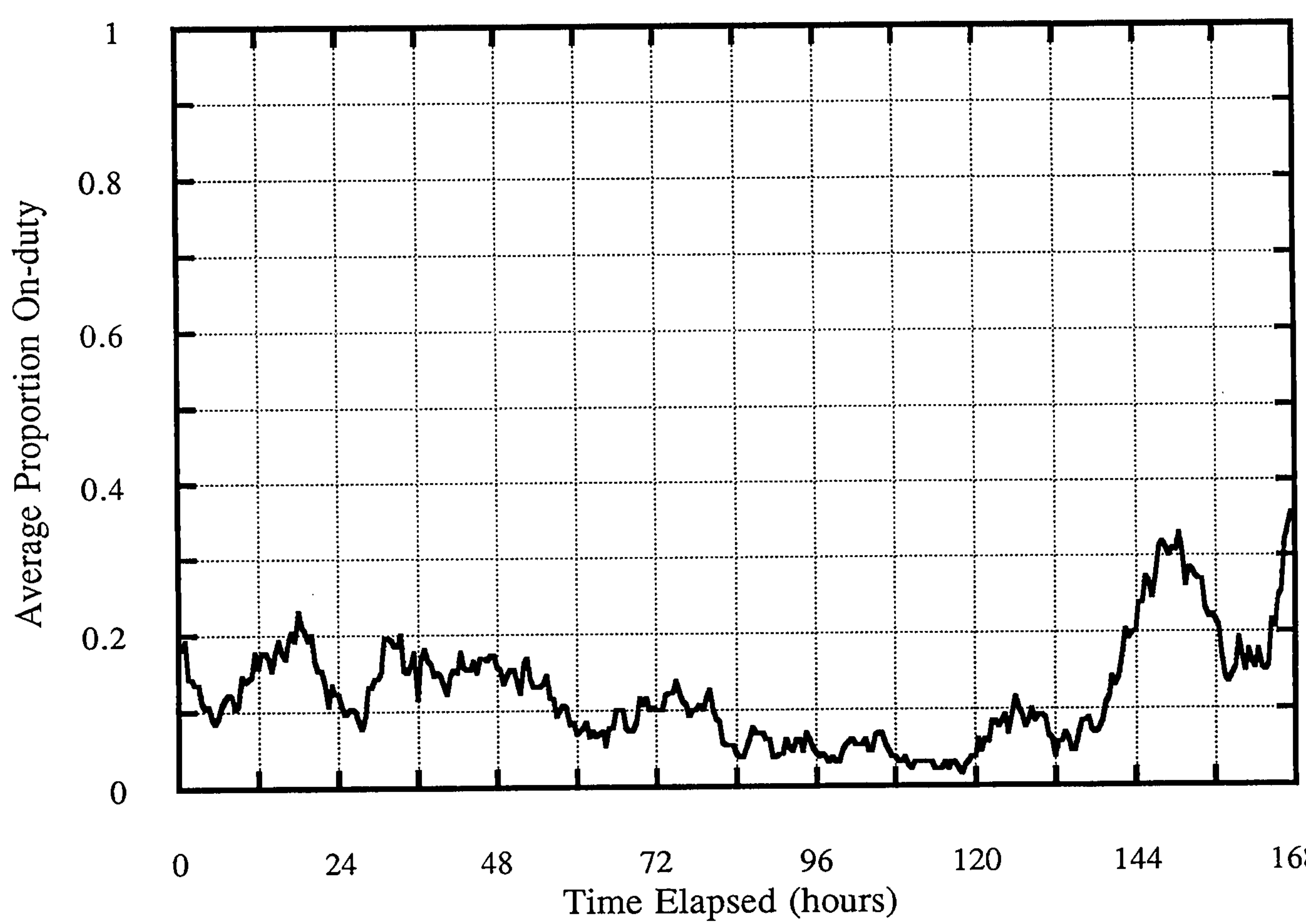
Pattern 4 Used for Modeling



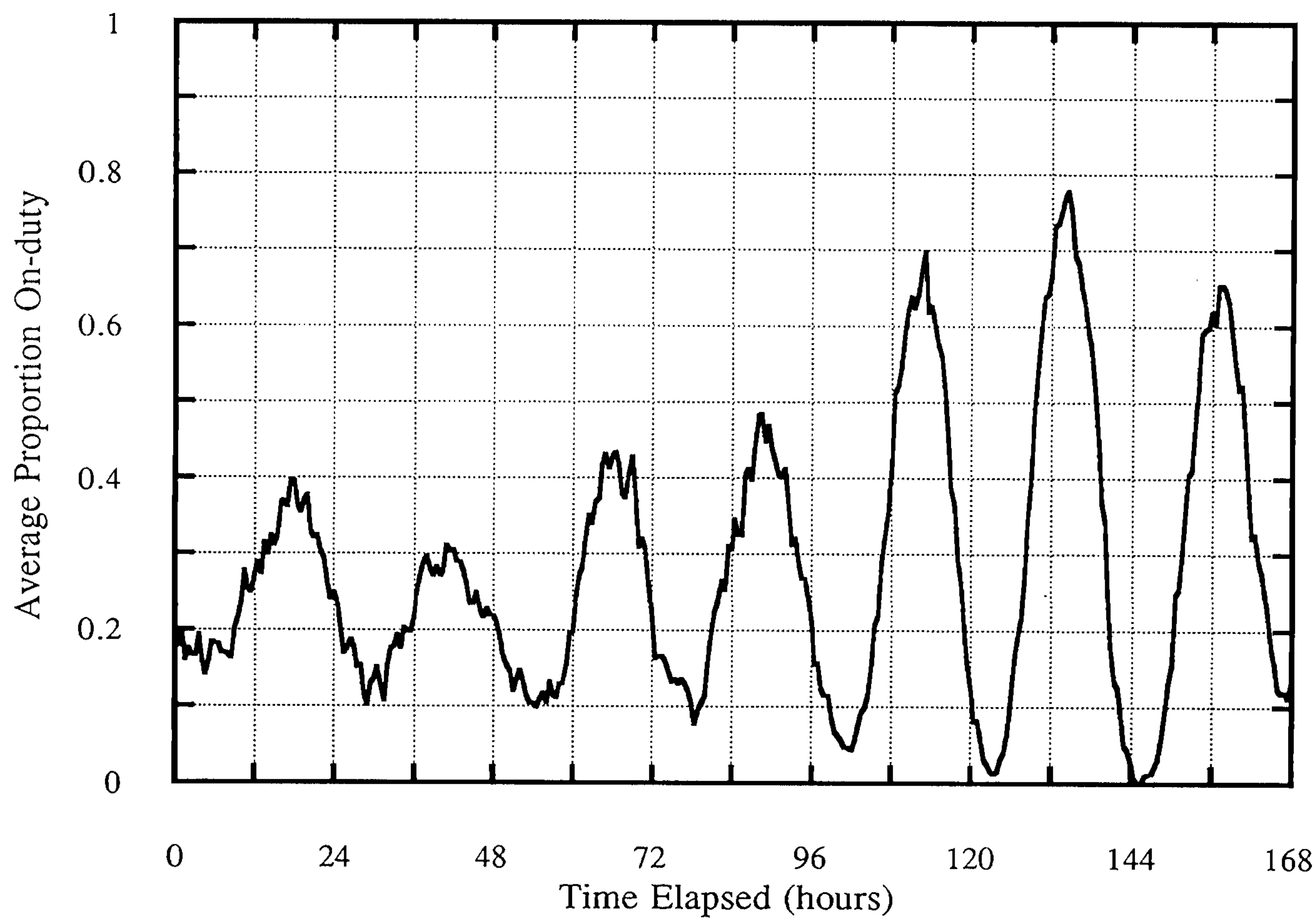
Pattern 5 Used for Modeling

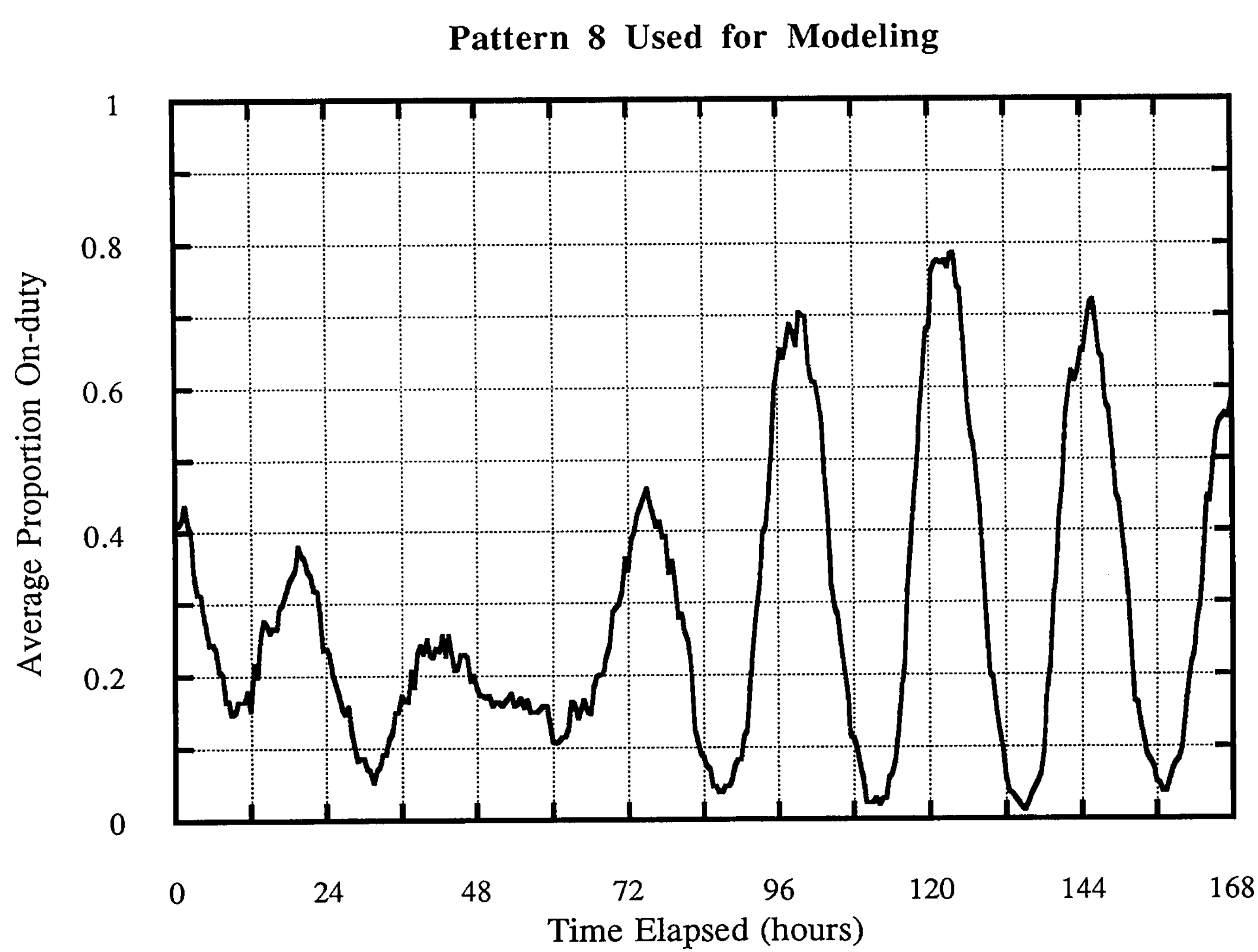


Pattern 6 Used for Modeling



Pattern 7 Used for Modeling




Pattern 8 Used for Modeling
Average Proportion On-duty
0
0.2
0.4
0.6
0.8
1
0
24
48
72
96
120
144
168
Time Elapsed (hours)

Pattern 9 Used for Modeling