Accid. Anal. & Prev. Vol. 23, No. 4, pp. 239–255, 1991
Printed in Great Britain.

0001-4575/91 \$3.00 + .00

ANALYSIS OF THE FREQUENCY AND DURATION OF FREEWAY ACCIDENTS IN SEATTLE

BRYAN JONES, LESTER JANSSEN, and FRED MANNERING
121 More Hall, FX–10, University of Washington, Seattle, WA 98195, U.S.A.

(*Received* 16 *October* 1989; *in revised form* 24 *August* 1990)

Abstract—Accidents occurring on congested urban freeways can have enormous impacts in terms of lost commuter time. This paper presents an appropriate statistical analysis of urban freeway accident frequency and duration and discusses how this analysis can be used to guide management strategies that seek to reduce the traffic-related impacts of accidents. The findings and demonstration of analysis procedures should be of considerable value to ongoing and future studies in this area.

INTRODUCTION

Persistent recurring traffic congestion is a problem facing virtually every major metropolitan area in the United States. With expectations of continued growth in traffic demand and relatively little money available for capital expansion of highway networks, effective mitigation of traffic congestion will continue to be an important area of national research focus. Within this context, perhaps one of the least understood and most disruptive phenomena facing traffic flow management is the mitigation of traffic congestion resulting from vehicle accidents. The traffic congestion costs associated with accidents, measured in terms of additional travel hours, can be quite substantial, as shown in a recent Seattle area study that found that each additional minute of accident duration can cost over \$2,000 in lost travel time (Garrison and Mannering 1990). This translates into estimated annual travel time losses in excess of \$250 million in Seattle alone. National recognition of the problem of accident-induced congestion, particularly its occurrence during peak traffic periods, has been building rapidly in recent years, but because of the complexity and random nature of accidents, truly effective accident management remains a much sought goal.

Perhaps the most important element in the development of management strategies that seek to reduce the traffic-related impacts of accidents is a thorough understanding of the factors affecting the frequency of accidents, as well as their duration (i.e. the time between their occurrence and the restoration of full roadway capacity). Such information is vital to the allocation of personnel and equipment both spatially and temporally. Moreover, a rigorous ongoing statistical assessment of accident frequency and duration can serve both as a measure of accident management effectiveness and as a basis for developing new strategies and/or reallocating existing management resources.

The objective of this paper is to demonstrate the usefulness of appropriate statistical techniques in the study of accident frequency and duration and ultimately in the formulation and assessment of accident management strategies. To achieve this objective, an accident frequency and duration study was undertaken in the Seattle metropolitan area. Seattle is a particularly interesting region for this type of study because of its highly congested urban freeway system, its accident mitigation program, and regional recognition of the need to improve accident management. As will be shown, the analysis methods and procedures demonstrated in this paper have important implications for accident management strategies in general and can be used as a basis for studies in other metropolitan areas.

The paper begins with a description of the study area and a discussion of the accident management strategies currently used in Seattle. A brief description of the data follows, including some intricacies of the data collection process. Next, the statistical techniques used to estimate accident frequency and duration are presented. These are followed by a discussion of accident frequency and duration model estimation results. Finally, the paper closes with a summary and some concluding remarks.

STUDY AREA AND EXISTING ACCIDENT MANAGEMENT

The study area is shown in Fig. 1. Portions of two routes were analyzed in this area: Interstate 5 and State Route 520 (SR 520). Both of these routes vary considerably in average congestion and in geometrics. For example, SR 520 has no shoulders and includes a floating bridge with 10-foot lanes. As a consequence, the traffic impacts of accidents on this route tend to be quite severe. In contrast, I–5 has 11- or 12-foot lanes and a shoulder for most of the portions of the study area. The factors affecting accident frequency and duration on I–5 are quite different from those on SR 520. However, the impact of a lane blocking accident on I–5 is still very high because the highway is now near capacity, and an accident creates additional weaving in an area already noted for weaving problems.

To account for the differing geometrics on the two routes, since geometrics may affect both the frequency and duration of incidents, the routes were subdivided into six roughly homogeneous zones, as illustrated in Fig. 1. Zones 1 and 2 were on SR 520, and Zones 3 through 6 were on I–5. These zones formed the basis for the accident data collection and analysis.

A number of resources are used to mitigate the traffic-related impacts of vehicular accidents in the 20-mile study area. For example, the entire study area is under 24-hour surveillance (by closed circuit television) at a transportation system system management center (TSMC). This surveillance system enables the Washington State Department of Transportation (WSDOT) to detect accidents fairly quickly (within an average of five minutes after an accident's occurrence during the system's operating hours). In addition, the Washington State Patrol has a well organized accident management effort, and troopers arrive at the scene of an accident in an average of just over four minutes after having been notified, regardless of time of day. The Seattle area's accident management effort also includes dedicated tow trucks in Zone 2 (see Fig. 1), which are in position during peak hours (6:00–9:00 A.M. and 2:00–7:00 P.M.) awaiting accident occurrence. A single accident equipment storage site (Zone 1), which contains materials (e.g. traffic cones, cleanup materials) that the State Patrol can use to mitigate accident traffic impacts, is also available. Unfortunately, the use of this facility has been limited for a number of reasons, including (i) lack of trooper awareness of its existence and (ii) communication limitations that often hamper the selection of proper materials before an accident site has been visually inspected.

Although current accident management in the 20-mile study area has been effective in reducing the traffic-related impacts of accidents, the WSDOT generally recognizes that additional improvement is needed. WSDOT views improved accident management as a relatively inexpensive approach to relieving traffic congestion on Seattle's highway network, and one that will generate little, if any, public opposition (a rarity in the congestion-mitigation field). The forthcoming data collection and analysis was conducted to provide direction for such improvement.

DATA DESCRIPTION

The data collection effort was designed to study the factors that affect accident frequency (number of occurrences) and duration (defined as amount of time between the time the officer receives a report of an accident and the time he or she leaves the scene of the accident).* Accident frequency data were derived from two sources: (i)

*This definition of duration was used because it provided for the most accurate times. The obvious definition of duration would be from the moment the accident occurred, but these times often are inconsistently reported by somewhat disoriented accident participants. Another possibility is to use the time at which the road was cleared to denote the end of the accident's duration. This option was rejected on the grounds that officers' definition of "cleared" is open to interpretation, and the use of the "cleared" time ignores the disruptive effect that the presence of the officer and possible vehicle debris has on traffic flow.

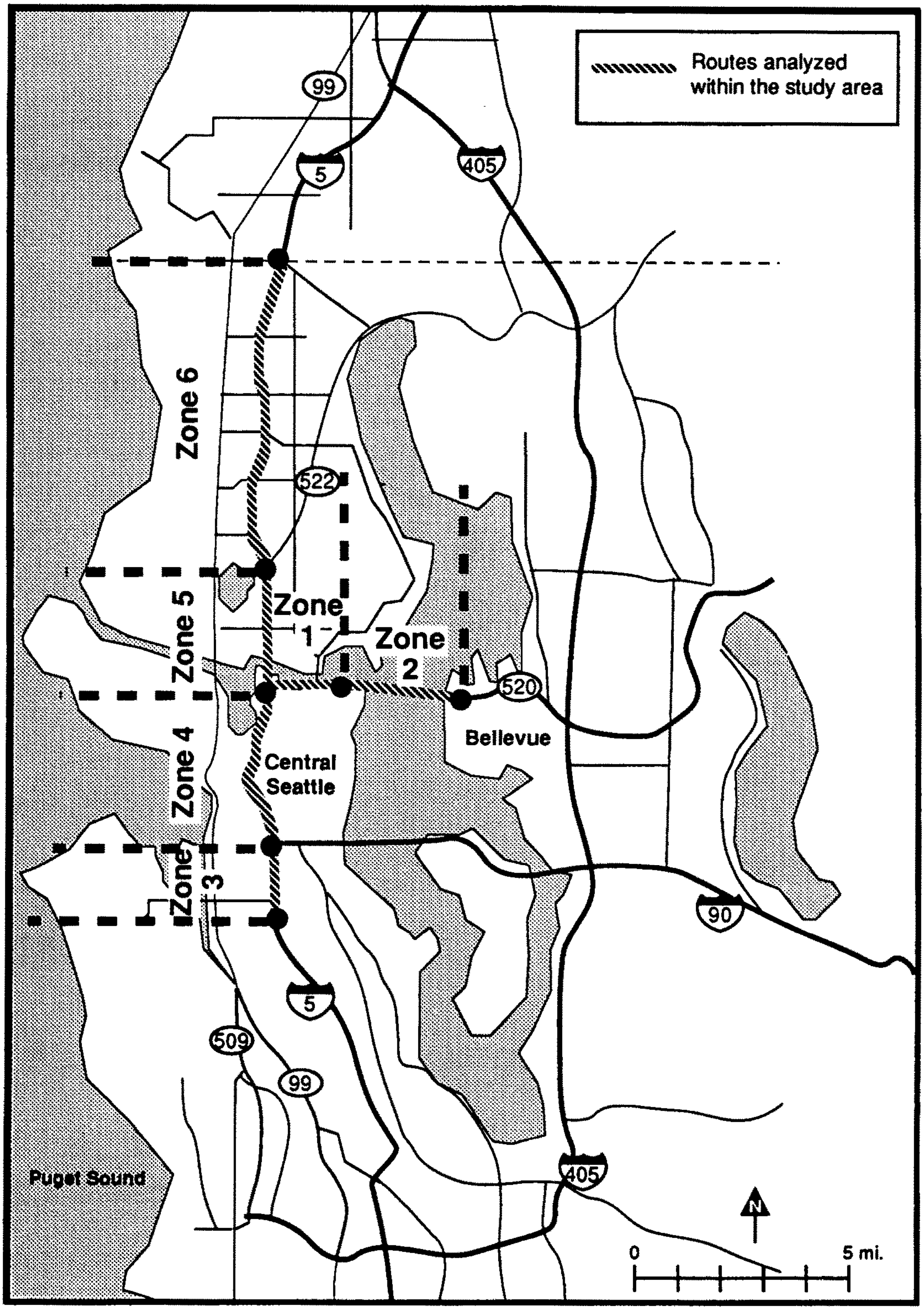

Fig. 1. Study area with incident location zones.

Washington state accident records and (ii) special events information. The data were collected from all accidents that occurred in the six study zones from April 1987 to March 1989. In all, 5,637 accident reports were obtained from the state in computer-ready form, and appropriate special event information was added to these data (see Table 1 for a summary of accident report data). Figures 2 and 3 provide a summary of these accident frequency data. The average number of accidents per day by zone showed that Zones

Table 1. Variables available from accident reports

Year	Year of accident
Month	Month of accident
Day	Day of Accident
Day of Week	Indicator for day of week Monday to Sunday
Hour	Hour accident report started
Minute	Minute accident report started
Sign Route	State route highway designation number
SR Milepost	State route mile post
Accident Sev	Accident severity index; property only, injury accident, or fatality
N. Injured	Number of persons injured in the accident
N. Fatal	Number of persons killed in the accident
Light	Indicator for illumination level at accident site: daylight, dawn, dusk, dark (with and without street lights, and other)
Collision Type	Code for various possible collision types including pedestrian/vehicle, vehicle/vehicle, parked vehicle and others kinds
Object Struck	Kind of object struck, if any (e.g. light standard)
M. Sev.Inj.	The most severe injury caused by the accident (no injury, fatal, disabling, non-disabling, possible, unknown)
N. Veh.	Number of vehicles involved in the accident
P.Dam.$	Property Damage measured in dollars

Table 1 (Continued)

R. Char.	Roadway character - grades and curves
L. Char.	Location character - codes for various intersections, under and over passes and other facilities
R. Sur.	Road surface character: not stated, dry, wet, snow, ice, other
Weather	Weather at the accident site:clear/cloudy, rain, snow, fog, or other
Res. Prox.	Residence proximity of involved drivers: within 15 miles, elsewhere in state, or out of state
Sobriety	Sobriety of the drivers in the accident: 7 codes for had been drinking — ability impaired to had not been drinking
A. Sev.	Alcohol severity: drunkest driver involved in accident
Con.Circ	24 codes indicating different possible RCW violations or indicating no violation
D.V.Act	Driver Vehicle Action: codes indicating evasive or non evasive actions taken by the involved drivers
Veh. typ.	Vehicle type: vehicle type code
Age	Age for each of the involved drivers
Haz. Mat.	Kind of hazardous material involved, if any
Fuel	Fuel Spill (yes/no)
Fire	Fire Resulted (yes/no)

4 and 6 had the highest frequencies (see Fig. 2), but when roadway length was accounted for, Zones 5 and 1 proved to be the most dangerous, as measured by the number of accidents per mile (see Fig. 3).

The primary source of accident duration data was the State Patrol dispatch records over the April 1987 to March 1989 period. The researchers acquired these data by tediously sifting through files of dispatcher cards and computer-coding the needed information (see Table 2). Since the forthcoming analysis of accident duration required information available from the accident reports as well as from the dispatch data, it was

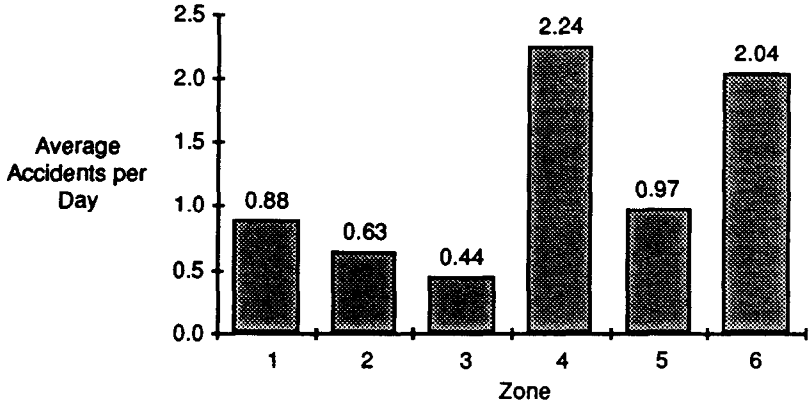

Fig. 2. Accidents per day by zone.

necessary to match the two data sets. Of the 5,637 accident reports, only 2,156 could be matched with corresponding dispatch data. As with the frequency data, appropriate special event information was also added to these data. Figure 4 gives the distribution of accident durations (for all zones) suggested by these matched data. Note that the shape of this distribution was approximately normal.

ESTIMATION METHODS: ACCIDENT FREQUENCY

To assess the frequency of accident occurrence, an appropriate statistical modeling technique is needed. Within this context, a Poisson distribution is a reasonable description of the number of traffic accidents in a given day. The Poisson regression model can effectively overcome the problems caused by discrete and non-negative values of observations that would be found in normal linear regression analysis (Mannering 1989). The Poisson distribution has previously been used in diverse count data applications that have ranged from trip delay frequency (Mannering and Hamed 1990) to beverage consumption frequency (Mullahy 1986). The Poisson regression has also been previously applied to accident frequency (for example, Jovanis and Chang 1987), for which the suitability of the technique has been demonstrated theoretically and empirically. The Poisson model is as follows:

$$P(n) = (\exp(-\lambda)\lambda^n)/n! \tag{1}$$

where $P(n)$ is the probability of having n accidents per day, and λ is the Poisson parameter, which will be some estimable function of the independent variables. Herein, the

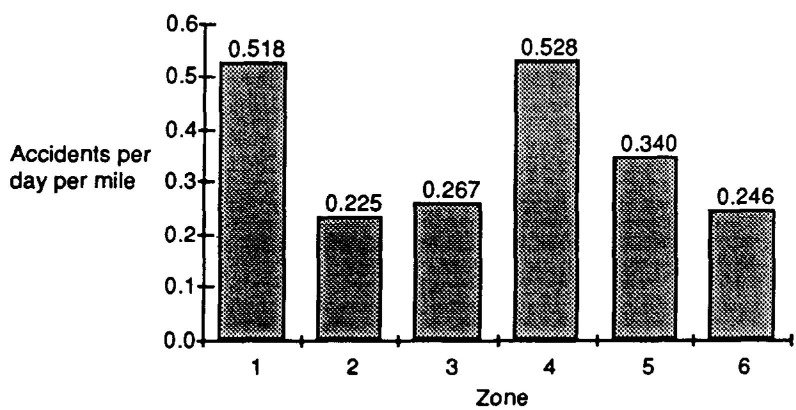

Fig. 3. Accidents per day per mile, by zone.

Table 2. Data collected from dispatch logs

Year	Year of accident
Date	Month and day of the accident
DOW	Day of week: Monday to Sunday
TOD	Time of day: one of eight 3 hour time slots
Rcvd	Dispatcher received call for assistance
Enroute	Time Trooper was en route to accident scene
Atscene	Time Trooper arrived at the accident scene
Roadcl	Time road was cleared, if it had been blocked
Troopcl	Time Trooper cleared the scene of the accident
Tow	Number of tow trucks called to accident site
Amb	Number of ambulances called to the accident site
Other	Number of other emergency vehicles called to the accident site (e.g. Fire Department)
Exp	Indicates that accident occurred on the express lanes
Loc. C	Location Code: code for the cross streets on I-5 and SR 520 within our study area
Dir	Direction of travel
Lane	Lanes(s) involved
N.Veh	Number of vehicles involved
N.L.Block	Number of lanes blocked
Inj	Number of injuries
Ftl	Fatality accident (yes/no)
T/B	Truck or bus involved in the accident

Poisson parameter is defined as follows:

$$\log \lambda = \boldsymbol{\beta} \mathbf{X} \tag{2}$$

where $\boldsymbol{\beta}$ is a vector of estimable parameters and $\mathbf{X}$ is a vector of commuting and other characteristics for the day. The object of the Poisson analysis is to estimate the vector, $\boldsymbol{\beta}$, thereby providing an estimate of the natural log of the mean number of accidents

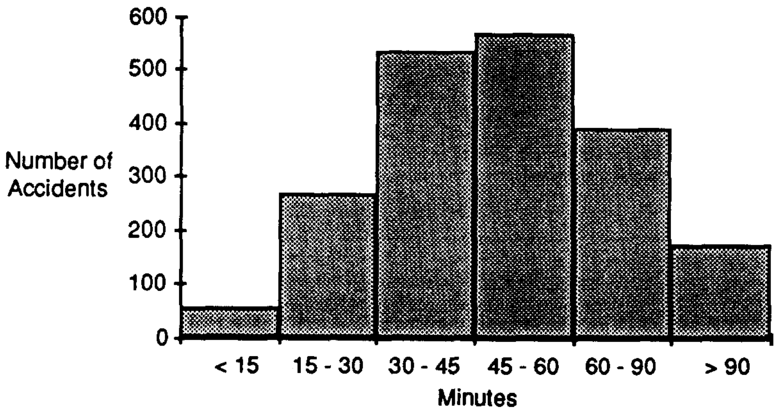

Fig. 4. Distribution of durations.

per day. This estimate is achieved with standard maximum likelihood procedures (Mannering 1989). Finally, note that, unlike standard least squares regression analysis, λ is a deterministic function of $\mathbf{X}$, with the randomness coming from the probability specification for n.

ESTIMATION METHODS: ACCIDENT DURATION

The length of time between the time a police officer receives a report of an accident until he/she leaves the accident scene is defined as the duration of the accident. In recent years, an entire body of literature has evolved to address the special problems associated with duration data. This literature has drawn heavily on statistical methods developed largely in industrial engineering where they are used to describe the useful lives of various machines, and in the biomedical sciences where they describe events such as the survival times of heart transplant recipients (Kiefer 1988).

The central concept in the study of duration data is not the unconditional probability of an event taking place (i.e. the probability of an accident lasting exactly ten minutes), but its conditional probability (i.e. the probability of an accident ending in the tenth minute given that it has lasted nine minutes). The special methods of duration analysis are a useful and convenient means of organizing, summarizing, and interpreting data for which a sequence of conditional probabilities is appealing.

Defining a duration precisely requires a time origin (a beginning), a time scale, and an end. The duration of a traffic accident, in our case, is the length of time as defined above. These durations are the dependent variables under study, and they are assumed to be affected by independent variables, such as the condition of the road surface at the accident site.

Central to the statistical estimation of a duration model is the hazard function. To define this function, note that the distribution of duration can be written as follows:

$$F(t) = Pr(T < t) \tag{3}$$

This equation specifies the probability that a random variable, T, is less than some specified value, t. In our case it is the probability that an accident has a duration less than t. Equation 3 also suggests a survivor function:

$$S(t) = Pr(T \geq t) = 1 - F(t), \tag{4}$$

which gives the probability that an accident has a duration greater than or equal to t. The corresponding density ($f(t)$) and hazard function ($h(t)$) are as follows:

$$f(t) = dF(t)/dt \tag{5}$$

$$h(t) = f(t)/S(t) \tag{6}$$

where $h(t)$ is roughly interpreted as the rate at which accident durations will end at time t, given that they have lasted for t minutes.

The hazard function provides a convenient definition of duration dependence. If the value of the derivative $dh(t)/dt$ at some t^* is greater than 0, the hazard increases in duration, indicating that the probability that an accident duration will soon end increases with the length of the accident. If $dh(t)/dt$ is less than 0, the hazard decreases in duration, and the longer the accident lasts the less likely it is to end soon. Finally, if $dh(t)/dt$ equals 0, the hazard is constant, and the probability of an accident duration ending is independent of time.

To estimate a duration model based on the hazard function, some interaction between the duration and the explanatory variables thought to affect duration must be specified. To achieve this, two modeling approaches have enjoyed considerable popularity (Kalbfleisch and Prentice 1980): the proportional hazards model and the accelerated lifetime model. The proportional hazard specification is easily applied and simple to interpret. The effect of time regressors (explanatory variables) is to multiply some baseline hazard function by a scale factor. In equation form, this approach has a hazard as follows:

$$h(t, \boldsymbol{\beta}, \mathbf{X}, h_0) = h_0(t)y(\boldsymbol{\beta}, \mathbf{X}) \tag{7}$$

where t is time, $\mathbf{X}$ is a vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of estimatable parameters, $h_0(t)$ is the baseline hazard (i.e. the hazard at $\boldsymbol{\beta}\mathbf{X} = 0$), and $y(\boldsymbol{\beta}, \mathbf{X})$ is a scaling factor.

The accelerated lifetime model uses explanatory variables to rescale time directly. In this case the hazard is

$$h(t, \boldsymbol{\beta}, \mathbf{X}) = h_0[ty(\boldsymbol{\beta}, \mathbf{X})]\, y(\boldsymbol{\beta}, \mathbf{X}) \tag{8}$$

where all terms are as previously defined. For both proportional hazards and accelerated lifetime models, the scaling factor is usually of the following form:

$$y(\boldsymbol{\beta}, \mathbf{X}) = \exp(\boldsymbol{\beta}\mathbf{X}) \tag{9}$$

where the parameter vector $\boldsymbol{\beta}$ is estimable by maximum likelihood techniques (Kalbfleisch and Prentice 1980).

Given these approaches, the application to accident duration requires that two choices: (i) selection between proportional and accelerated lifetime models and (ii) selection of an appropriate failure distribution. In the general area of accident studies, Jovanis and Chang (1989) have used the proportional hazards approach to study the time until an accident for individual trips. Specifically, they have applied the well-known Cox semi-parametric proportional hazard (Cox and Oakes 1984). The key advantage to this model is that one need not specify a distribution of the hazard function. However, in the context of accident duration, theoretical and empirical evidence suggests the validity of specific hazard distributions. For example, Golob, Recker, and Leonard (1987) have given theoretical justification for accident durations to be viewed as log-normal distributed. This distribution of accident duration has been empirically supported by Golob et al. (1987), as well as Giuliano (1988).

To study the accident duration data, we selected the accelerated lifetime approach and assumed an appropriate distribution.* However, unlike earlier studies that found the log-normal to be suitable for accident duration analysis, empirically this study's data refuted the validity of the log-normal in favor of the "wider" probability tails provided by the log-logistic distribution.** It should be noted, in support of the earlier work of

*Given the nature of our data (i.e. no censoring and the lack of time dependent covariates), there is no strong theoretical or empirical reason for selecting one of the two methods (i.e. proportional hazards, accelerated lifetime) over the other.

**In fact, Gamma, Weibull, and log-normal were all found to be empirically weaker than the log-logistic distribution.

Golob et al. (1987), that the log-logistic is a close approximation of the log-normal and is frequently used in duration studies as such (Kalbfleisch and Prentice 1980).

Some important characteristics of the log-logistic distribution are worthy of note, as they will prove useful in the interpretation of the model estimation results. With shift parameter $\gamma > 0$ and scale parameter $\rho > 0$, the log-logistic hazard function is as follows:

$$h(t) = \gamma\rho(\gamma t)^{\rho-1}/[1 + (\gamma t)^{\rho}] \tag{10}$$

This equation suggests that the hazard is monotone decreasing from infinity if $\rho < 1$ and is monotone decreasing from γ if $\rho = 1$. If $\rho > 1$, the hazard first increases from 0 to a maximum at $t = (\rho - 1)^{1/\rho}/\gamma$ and then decreases monotonically toward 0. As will be shown later, the log-logistic scale parameter, ρ, provides an important empirical interpretation.

ACCIDENT FREQUENCY MODELS

Accidents per day were modeled with a Poisson regression, as described earlier. Six models (one per zone, see Fig. 1) were developed to estimate accident frequency and to identify the characteristics peculiar to a specific day that might increase or decrease the number of expected accidents. The discussion begins with a comparison of these zonal models. The models' independent variables are discussed by variable class: seasonal effects, weekly trends, special events, and environmental factors. A discussion of the summary statistics follows. Table 3 shows each Poisson regression model, its variables, coefficients, t-statistics, and summary statistics.*

It is important to note that many of the variables included in Table 3 were acting, at least in part, as proxies for traffic volume data. Although traffic data were available from the magnetic loop detectors operated in the zones, the reliability of the data was suspect because of inoperable magnetic loop detectors in several key cases. Subsequently, month and day-of-week indicator variables, which have an immediate intuitive appeal, were used to implicitly capture traffic variability.

Seasonal effects

Seasonal effects are measured with month indicator variables. These indicator variables implicitly account for factors such as seasonal variations in traffic, the interaction of unmeasured weather effects and roadway geometrics,** and fluctuations in travelers' origin-destination patterns. The combination of these factors result in month-indicator coefficients that vary significantly from month to month and even from zone to zone within the same month. It is interesting to note that statistical tests bore out the temporal stability of these month-indicator coefficients (i.e. over the April 1987 to March 1989 data period). Thus, there appeared to be consistently strong and persistent seasonal effects.

Weekly trends

Day-of-week indicator variables are intended to exclusively capture variations in traffic volume. Each day is modeled with respect to Sunday, which is implicitly set to zero. Thus, since Sunday volumes and corresponding accident exposure levels are low, positive values for all day-of-week variables are expected. In general, the day-of-week indicators were highly significant and strongly positive.

Special events

The effect of planned special events on accident frequency was explored by including indicator variables for scheduled game days of local sports teams: the Seahawks football

*For specific zonal models, dashes indicate the exclusion of a variable on the grounds that it was statistically insignificant. As a rule, the 85% confidence level was used to determine statistical insignificance.

**As will be shown shortly, measured weather effects include road surface conditions and the presence of rain. Unmeasured weather effects include factors such as sun angles, wind, and temperature.

Table 3. Comparison of accident frequency models (t-statistics are given in parentheses)

Variable		Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
Intercept		-0.1310 (-6.662)	-1.1410 (-4.923)	-1.6457 (-6.743)	-0.3113 (-2.784)	-1.1320 (-6076)	-0.4225 (-3.644)
February	(1 if February, 0 otherwise)	0.4760 (3.469)	0.5606 (3.437)	— —	-0.1720 (-1.649)	— —	0.1277 (1.334)
March	(1 if March, 0 otherwise)	0.4602 (3.473)	0.4697 (2.862)	— —	— —	— —	0.1617 (1.784)
April	(1 if April, 0 otherwise)	0.3653 (2.652)	0.2267 (1.277)	— —	0.1464 (1.653)	— —	— —
May	(1 if May, 0 otherwise)	0.5972 (4.736)	0.5735 (3.703)	— —	-0.1316 (-1.310)	— —	— —
July	(1 if July, 0 otherwise)	-0.2801 (-1.547)	— —	— —	— —	0.2973 (2.279)	0.2158 (2.328)
August	(1 if August, 0 otherwise)	— —	0.3595 (2.145)	0.3667 (1.941)	0.0913 (0.982)	-0.2385 (-1.455)	— —
October	(1 if October, 0 otherwise)	— —	— —	0.3149 (1.663)	0.1561 (1.793)	0.1943 (1.475)	— —
December	(1 if December, 0 otherwise)	— —	-0.2435 (-1.125)	0.3977 (2.259)	0.1246 (1.443)	0.1971 (1.500)	0.2382 (2.748)
Monday	(1 if Monday, 0 otherwise)	0.2805 (1.613)	0.3185 (1.391)	0.5275 (2.169)	0.5777 (5.072)	0.1660 (1.031)	0.6278 (5.629)
Tuesday	(1 if Tuesday, 0 otherwise)	0.5827 (3.553)	0.5448 (2.489)	0.7263 (3.070)	0.6830 (6.066)	0.2817 (1.778)	0.4970 (4.357)
Wednesday	(1 if Wednesday, 0 otherwise)	0.6739 (4.172)	0.9635 (4.650)	0.6740 (2.809)	0.7263 (6.494)	0.1678 (1.034)	0.5372 (4.770)
Thursday	(1 if Thursday, 0 otherwise)	0.6315 (3.892)	1.0890 (5.380)	0.4946 (2.008)	0.7847 (7.104)	0.2765 (1.747)	0.5895 (5.272)
Friday	(1 if Friday, 0 otherwise)	0.5139 (3.096)	0.6828 (3.216)	0.6865 (2.906)	1.0273 (9.668)	0.7259 (4.990)	0.8221 (7.588)
Saturday	(1 if Saturday, 0 otherwise)	0.2887 (1.062)	0.7861 (3.727)	0.4608 (1.861)	0.5542 (4.826)	0.4245 (2.670)	0.4384 (3.790)
Seahawks	(1 if Seahawks game, 0 otherwise)	— —	0.4601 (1.519)	0.5630) (1.921)	0.4866 (3.476)	0.3614 (1.671)	— —
Huskies	(1 if Huskies game, 0 otherwise)	0.3774 (1.270)	— —	— —	— —	0.8391 (4.155)	— —
Mariners	(1 if Mariners game, 0 otherwise)	0.1498 (1.464	0.1591 (1.309)	— —	— —	0.1577 (1.646)	— —
Sonics	(1 if Sonics game, 0 otherwise)	— —	0.309 (2.065)	0.2524 (1.463)	— —	— —	0.1051 (1.271)
Road Surface	(1 if wet surface, 0 otherwise)	0.2777 (3.741)	0.2421 (2.101)	0.1110 (0.982)	0.3317 (7.470)	0.2065 (2.438)	0.2974 (5.349)
Weather	(1 if raining, 0 otherwise)	— —	— —	— —	— —	0.2341 (2.825)	0.1244 (2.060)
Summary Statistics							
Number of Observations		731	731	731	731	731	731
Log Likelihood at Zero		-1034.4	-943.3	-811.3	-1933.0	-1027.6	-1738.3
Log Likelihood at Convergence		-985.5	-842.0	-651.6	-1427.2	-976.1	-1363.7
RHO Squared		0.047	0.1073	0.197	0.2617	0.0501	0.2155

team and the Mariners baseball team (the stadium is near Zone 3), the Sonics basketball team (stadium is near Zone 4), and the University of Washington Husky football team (stadium is near Zone 1). Lower than average t-statistics might have been expected for the football game days, since all Huskies games were played on a Saturday and all Seahawks games were played on Sunday. However, a Pearson's correlation coefficient analysis did not show that Saturday and Sunday were highly correlated to Huskies and Seahawks games, and the high t-statistics for both variables in zones near the game sites did bear this out. Proximity to the game site and presence of important weaving sections seemed to have had the greatest effect. The Seahawks played their games in a stadium near Zone 3. On Interstate 5, Zones 3, 4, and 5 showed significant positive effects of Seahawk games on accident frequency. Zone 6 was too far removed from the facility to show any significant effect. Interestingly, Zone 2 on State Route 520, not Zone 1, had a significant t-statistic for Seahawk games, which was likely the outgrowth of weaving areas in this zone.

Husky football games had a significant effect in Zone 1, which is adjacent to the game site, and Zone 5, which accommodated much of the game day parking overflow.

Considering the comparatively low attendance at the Mariners baseball games (played in the same stadium used by the Seahawks), the increase in expected number of accidents for those game days in Zones 1 and 2 (the SR 520 section) and in Zone 5 was surprising. Each of these zones had comparable magnitudes for their coefficients and t-statistics. Also puzzling was that Zones 3 and 4 (near the stadium, where Mariners games are played) did not show any significant effects. This omission was a result of game traffic diversion to arterials and other surface streets, since many of the games were scheduled for weekdays near the evening rush hour.

The Seattle Sonics basketball team played their games in a stadium near Zone 4. Their games were scheduled for both weekdays and for weekends. Zones 2, 3, and 6 had significant coefficients for Sonics game days, although the coefficient of Zone 6 was marginally significant. The significance of these games only for zones further away from the game site was the result of route diversion to arterials as spectators neared the game site.

Environmental factors

Two environmental factors were included in the model: road surface character ("wet surface") and weather (raining vs. all other weather types). The rain indicator showed surprisingly little effect on accident frequency. Only Zones 5 and 6 had significant results for rain. Zones 5 and 6 included northbound merges into the fast lanes and other complex weaving maneuvers that contributed to increased accident rates. The road surface indicator was significant for all zones and was derived from the accident reports, from which a daily average was determined. The amount of variation within each day's reported road surface was very low.

Overall model statistics

The total number of observations for each model was 731, one observation for each day of the two-year study period. Rho squareds are given at the bottom of Table 3. Zones 3, 4, and 6 had relatively high rho squares, indicating that an acceptable level of the dependent variable variation (i.e. accidents per day for each zone) was explained. The Zone 2 model performed less well, and Zones 1 and 5 had a considerable amount of unexplained variance. Still, significant variables for accident frequency were identified in all zones, and their relative magnitudes indicated how accident frequency was affected.

ACCIDENT DURATION MODELS

The dependent variable in the accident duration model was measured from the time the police officer was notified of the accident to the time the officer left the accident scene.* Again, models were estimated for each analysis zone, and the coefficient estimates are presented in Table 4. The classes of variables used were different from those used for the frequency models, mostly because the data pertained to specific events.

*This measurement implied that the duration definition consisted of two parts: (i) the response time of troopers and (ii) time to clear the accident. Modeling these times separately was not undertaken because of the relatively low average value of response times (four minutes), compared clearing times, and the high negative correlation between response and clearing times (i.e. severe accidents, which had long clearing times, tended to have short response times), which would have substantially complicated the statistical form of the duration models. Consideration of the time to detection was not undertaken, since the reliability of the reported time-of-accident occurrence data was questionable.

Table 4. Comparison of log-logistic duration models for each zone (t-statistics are given in parentheses)

Variable		Zone 1	Zone 2	Zone 4	Zone 5	Zone 6
Intercept		3.7367 (14.95)	3.6403 (33.06)	3.6170 (45.78)	3.8742 (28.68)	3.7068 (28.51)
April	(1 if April, 0 otherwise)	— —	— —	— —	0.1302 (1.56)	— —
July	(1 if July, 0 otherwise)	— —	— —	— —	0.1717 (2.21)	— —
August	(1 if August, 0 otherwise)	0.1953 (1.74)	— —	— —	— —	— —
September	(1 if September, 0 otherwise)	— —	— —	— —	— —	0.0942 (1.64)
October	(1 if October, 0 otherwise)	— —	— —	— —	— —	-0.1765 (-2.90)
November	(1 if November, 0 otherwise)	0.2663 (1.93)	— —	— —	0.0849 (1.20)	— —
Accidents per Day		0.0106 (1.77)	— —	— —	— —	— —
Rush	(1 if peak hours, 0 otherwise)	-0.2281 (-2.58)	-0.0810 (-1.12)	-0.0825 (-1.62)	— —	-0.1860 (-4.73)
Dark	(1 if night accident, 0 otherwise)	— —	— —	0.1343 (3.13)	0.1663 (3.44)	— —
Northbound	(1 if northbound lanes, 0 otherwise)	— —	— —	0.0500 (1.36)	— —	— —
Funtime	(1 if Friday or Saturday night, 0 otherwise)	— —	— —	0.0795 (1.25)	— —	-0.1553 (-2.75)
Husky	(1 if Huskies game, 0 otherwise)	-0.2730 (-1.30)	— —	— —	— —	— —
Alcohol	(1 if alcohol involved, 0 otherwise)	— —	-0.1155 (-1.52)	-0.0772 (-1.50)	-0.1321 (-2.31)	— —
Young Driver	(1 if driver less than 65, 0 otherwise)	— —	— —	— —	— —	-0.0721 (-1.14)
Truck or Bus, Involved (1 if yes, 0 if no)		— —	— —	0.2318 (3.64)	0.2102 (3.21)	0.1785 (3.32)
Vehicles Only	(1 if only vehicles involved, 0 otherwise)	— —	— —	— —	— —	0.0796 (1.94)
Number of Vehicles, in accident		— —	0.0688 (2.25)	0.0412 (1.83)	0.1337 (5.74)	— —
Property Damage (Thou. $)		— —	0.0213 (2.130)	0.0164 (1.64)	— —	0.0273 (3.90)
Property Damage Only	(1 if only property damage, 0 otherwise)	— —	— —	— —	-0.1856 (-4.15)	— —
Injury Accident	(1 if injury occurred, 0 otherwise)	— —	— —	— —	— —	0.1003 (1.78)
Number of Injuries, in accident		0.0953 (2.58)	— —	— —	— —	— —
Number of Lanes Blocked		— —	— —	0.0869 (3.02)	— —	0.0610 (2.33)
Log-Logistic Scale Parameter		0.2313 (16.17)	0.0688 (2.25)	0.0261 (2.93)	0.2343 (22.10)	0.2631 (31.70)
Summary Statistics						
Number of Observations		183	267	622	346	716
Log-Likelihood at Convergence		-98.2	-162.5	-420.5	-194.2	-487.6

New classes included driver characteristics and accident severity measures. The "weekly trends" class was dropped and replaced by "daily variations," which included the "rush hour variable," "accidents per day," "dark," and "funtime" variables. The rush-hour variable indicated whether an accident occurred during the morning or evening rush hours (6–9 A.M. and 3–6 P.M.) in the corresponding rush-hour direction. The "accidents per day" variable was a proxy measure of the State Patrol workload in the study area. The "dark" variable indicated accidents that occurred after sunset, and the "funtime" variable indicated accidents that occurred between 7 P.M. and midnight on Fridays and Saturdays, traditional recreational periods.

Seasonal effects, daily variations, special events, driver and vehicle characteristics, and accident severity measures are discussed in turn below. Note that Zone 3 was left out of the duration analysis because of an insufficient sample size.

Seasonal effects

As with the accident frequency models, the month-indicator variables, used to assess seasonal effects, implicitly capture a host of factors including traffic variations and interactions with roadway geometrics. Overall, the duration of accidents was not consistently affected by seasonal variations to the extent that frequency of accidents was. However, some months did have significant effects for particular zones.

Daily variations

Accidents per day were significant only in Zone 1. As suggested earlier, this zone has very restrictive roadway geometrics with no shoulders, and additional accidents would have created more merging sections in this zone, making accident investigation and vehicle removal much more difficult and thus increasing overall accident duration.

The rush hour indicator showed a decrease in accident duration during the peak hours (6–9 A.M.. and 2–7 P.M.) in Zones 1, 2, 4, and 6. In Zone 2, the state DOT had tow trucks stationed during peak hours, and this service, a key advantage to vehicle clearing, was unavailable during the rest of the day. In general, the peak hours were recognized as problem times for traffic, and there seemed to be a conscious effort to reduce accident clearing times during these hours.

After dark, accidents were longer in Zones 4 and 5. In these zones, poorer sight distances and difficult investigation environments contributed to longer clearing times. Night accidents also tended to be more severe in these zones, since drivers may not have recognized hazards until too late.

As expected, the northbound accident indicator variable was significant in Zone 4 (Zones 1 and 2 run east and west). The northbound portion of Zone 4 has very restrictive geometrics, and, as a consequence, there are few places where an accident does not interfere with access or egress from ramps in this area and thus contribute to longer durations.

The "funtime" indicator was positive for Zone 4 and negative for Zone 6. These coefficient signs reflected personnel allocation at these times, which is high in Zone 6 (i.e. giving shorter durations) and comparatively low in Zone 4.

Special events

Only the Husky indicator was found significant in any of the zones. A highly organized, cooperative effort of local police, the University of Washington, and transit agencies in Special Event Incident Management produced a negative coefficient in Zone 1, the section of highway adjacent to the game site. More accidents were expected on game days, as seen in the frequency models, but the level of police presence and excellent response drove down the time necessary to clear accidents and complete investigations. This anticipated response demonstrates the potential effectiveness of well-demonstrated accident management programs.

Driver and vehicle characteristics

Approximately one-third of the drivers involved in an accident in the study area had been drinking alcoholic beverages. The impact of these drivers on accident duration was not an increase but a decrease in the length of time to clear the accident. This result suggests a higher level of trooper response to such accidents, which typically results in shorter accident clearing times. In fact, conversations with troopers indicated that multiple-car response is typical for accidents involving alcohol.

Drivers under 65 years old tended to have shorter accident durations for Zone 6. In terms of size, the other zones were relatively small compared to Zone 6, and the high traffic volumes put everyone at about the same risk. Farther from central Seattle, the effect of individual driver characteristics became more pronounced.

Truck or bus accidents had strongly significant coefficients for the Interstate 5 zones. State Route 520 (Zones 1 and 2) did not show a significant effect, which was a result of comparatively low truck volumes.

Accidents involving only vehicles were longer than accidents involving creatures such as dogs or ducks. Zone 6 had more greenbelt than other zones, so accidents involving animals could be adequately analyzed (i.e. for a significant vehicle-only variable). The other zones had an insufficient number of animal accidents.

The number of vehicles involved in an accident raised the duration of the accident significantly in Zones 2, 4, and 5. The poor geometrics of these zones (e.g. narrow shoulders and weaving areas) are problematic, particularly when the number of involved vehicles is high (i.e. clearing wrecked vehicles is often a difficult and time-consuming task).

Accident severity measures

Property damage, measured in thousands of dollars, was significant in indicating longer accidents for Zones 2, 4, and 6. Zone 1 did not show a significant effect attributable to property damage, perhaps because of the low variance of damage in this zone. The t-statistics were relatively high for this variable (the average property damage per accident was \$2,415).

"Property damage only" was an indicator variable that separated property damage from injury and fatal accidents. Zone 5 did not show a significant effect on duration by dollars of property damage, but it did indicate that accidents involving only property damage were expected to have a shorter duration than accidents involving injuries and fatalities. This variable was not significant in any other zone. The significance of this variable was attributable to the fact that this zone had a higher mix of injuries, property damage, and fatalities than other zones.

Injury accidents were significant for Zone 6 (i.e. if injuries were involved). The number of persons injured was significant for Zone 1, indicating an increasing duration as the number of injuries increased. In all, four different measures of accident severity were significant for different zones. For Zones 2, 3, and 5, severity was indicated significantly by the number of vehicles involved in the accident; in Zone 1, the significant severity measure was number of persons injured; for Zone 6 the significant severity measure was property damage only; for Zones 4 and 6, the number of lanes blocked was the significant severity measure.

Finally, for all models, the log-logistic scale parameter was less than 1, indicating that the hazard function was decreasing throughout. This means that the longer the accident lasted, the less likely it was to be cleared soon. This finding suggests that increasing severity is a problem at all severity levels, and not just at the most catastrophic accident extremes, as might be expected. If only the most severe accidents were problematic, the log-logistic scale would have been greater than 1, indicating an initially increasing hazard (i.e. less severe accidents were likely to end soon the longer they lasted) and then a decreasing hazard (i.e. only the most severe accidents were less likely to end soon the longer they lasted). Thus, as severe accidents become better managed, one would expect the scale parameter to exceed 1.

One qualification is necessary before this severity discussion is concluded. Recall that the accident duration data were drawn only from accidents that could be matched. As previously discussed, these matched accidents naturally tended to be more severe than accidents in general, because of more careful reporting. Thus, since the data were biased somewhat toward severe accidents, the log-logistic scale parameter should have been lower than it would have been if duration information had been available on all accidents. However, most likely this downward bias was not strong enough to negate the validity of the scale parameter discussion above.

APPLICATION OF FINDINGS

The results of this study are currently being used to assist in the ongoing development of Seattle's accident management system. The results provide management assistance in two important areas: (i) the allocation of resources relating to detection and response and (ii) on-site accident management/clearing procedures.

In terms of resource allocation, the Poisson model of accident frequency provides important direction as to where and when detection and response should be enhanced. For example, the general seasonal trends (as reflected by the month-indicator variables) provide information that can be used for longer-term planning, whereas the weekly variations (as reflected by daily-indicator variables) can assist in allocation by day. The coefficients for special events (sporting-game indicators) and environmental factors (road surface conditions and the rain indicator) provide valuable information on the likelihood of accidents at specific locations (zones).

The duration models provide information of a more provocative and suggestive nature in relation to accident management/clearing procedures. Some of the basic findings, such as problems with severe accidents (as suggested by the log-logistic scale parameter), come as no surprise. However, several of the estimated coefficients (such as the rush hour, Husky, and alcohol indicator coefficients) do provide insight as to which accident management procedures are really working. These coefficients raise important questions, such as, what exactly is being done during rush hours that results in a duration reduction? Can anything be learned from the observed duration reduction experienced during the University of Washington's Husky football games? Exactly what procedures are used to reduce duration when alcohol is involved and why are these procedures not applied in all cases? As the application of these findings proceeds, the researchers are discovering that the answers to these questions cannot be explained by a simple "allocation of resources." Instead, they have found weaknesses in several management procedures and an unfortunate perception among accident management personnel that accidents not occurring during rush hours and Husky games, and those not involving alcohol-impaired drivers, are somehow less important. They have also found the perception that once an accident's duration becomes long, less effort should be spent to clear it quickly since drivers are assumed to have found alternate routes. This perception goes a long way to explain the estimates of the log-logistic scale parameter.

SUMMARY AND CONCLUSIONS

This paper has demonstrated how appropriate multivariate statistical models of accident frequency and duration can be used to isolate key relationships between site and condition characteristics and the frequency and duration of accidents. Such relationships can serve as a guide to uncover the underlying strengths and weaknesses of existing accident management systems and provide important directions for management system improvement. The results of this analysis are currently aiding a major effort to improve Seattle's accident management system.

In terms of applying the methodology presented in this paper to other metropolitan areas, the key concern is the acquisition of an appropriate and accurate data set. In this regard, it is important to note that in the absence of data relating to roadway geometrics, specific accident management characteristics (e.g. trooper allocations), and traffic volumes, transferability among study zones within and between metropolitan areas is problematic.* It would be interesting to test the transferability of frequency and duration

*In our study, models were estimated over all zones and compared with the individual zonal models. Strong statistical justification was found for separating the estimates by zone for both frequency and duration models, indicating that the transferability among zones was not statistically possible.

models that explicitly included such detailed data, since proving transferability could greatly reduce the primary cost of data collection for accident frequency and duration studies.

Acknowledgment—The authors gratefully acknowledge the constructive and substantive comments provided by the two referees.

REFERENCES

Cox, D. R.; Oakes D. Analysis of survival data. New York: Chapman and Hall; 1988.

Garrison, D.; Mannering, F. Assessing the traffic impacts of freeway incidents and driver information. ITE Journal 60 (7):19–23; 1990.

Giuliano, G. Incident characteristics, frequency, and duration on a high volume urban freeway. Irvine, CA: Institute of Transportation Studies, University of California; 1988.

Golob, F.; Recker, W.; Leonard, J. D. An analysis of the severity and incident duration of truck-involved freeway accidents. Accid. Anal. Prev. 19:375–395; 1987.

Jovanis, P. P.; Chang, H. L. Modeling the relationship of accidents to miles traveled. Transp. Res. Rec. 1068:42–51; 1987.

Jovanis, P. P.; Chang, H. L. Disaggregate model of highway accident occurrence using survival theory. Accid. Anal. Prev. 21:445–458; 1989.

Kalbfleisch, J.; Prentice, L. The statistical analysis of failure time data. New York, NY: John Wiley & Sons; 1980.

Kiefer, N. M. Economic duration data and hazard functions. J. Economic Literature, 26(2):646–679, 1988.

Mannering, F. L. Poisson analysis of commuter flexibility in changing routes and departures times, Transp. Res. 23B:53–60; 1989.

Mannering, F. L.; Hamed, M. Occurrence, frequency, and duration of commuters' work-to-home departure delay. Transp. Res. 24B:99–110; 1990.

Mullahy, J. Specification and testing of some modified count data models. J. Econometrics 33(3):341–365; 1986.