USE OF A TELEPHONE SURVEY AND LOG-LINEAR MODELS TO EVALUATE A FEDERAL MOTOR VEHICLE SAFETY STANDARD

David Morganstein and John Burke, Westat, Inc.

INTRODUCTION

This paper discusses a telephone survey conducted by Westat, Inc. for the National Traffic Safety Administration (NHTSA), U. S. Department of Transportation. The survey was part of an evaluation undertaken by the NHTSA. The goal of the study was to obtain data on very low-speed car accidents for use in evaluating the effectiveness of Federal Motor Vehicle Safety Standard (FMVSS) 215 and Title I (damageability) Bumper Standards with regard to the reduction of property damage associated with low-speed autombile accidents.

The overall objective of the study was to collect and analyze data needed to evaluate the effectiveness of automobile bumpers in reducing property damage. The survey measured several indicators of effectiveness: the probability of damage; the number of components damaged; and the cost-to-repair damage. All of these measures of effectiveness were conditional on the occurrence of a low-speed eligible incident (that is, an accident which left the car drivable and was not reported either to the police or to an insurance company).

Survey Design

The survey design consisted of a combination prospective and retrospective national study of automobile drivers having low-speed collisions. The retrospective data were collected from drivers identified in a national telephone screening. Drivers indicating that they had an eligible incident during the prior six months were administered a questionnaire over the telephone regarding the incident. All drivers contacted were asked to continue participating in the survey by reporting in a followup interview all incidents occurring in the ensuing two-month, bounded period.

The initial identification and interviewing of respondents was completed by telephone from Westat's telephone facility in Rockville, Maryland. A screening instrument was administered to any household member 16 years of age or older. During the screening interview, the principal driver of each car in the household was identified and was asked whether or not the car had been in an incident of interest within the last six months. If it was, the interviewer determined the best person to interview about the incident.

Upon completion of the interview, the respondent was asked to participate during the two-month followup period. Immediately following the contact, respondents were sent a log in which they could record data on incidents that occurred during the followup period. One month after the initial contact, the respondents were sent a postcard reminder. For the followup interview, about 90 percent of the respondents were recontacted by telephone and the remaining 10 percent were recontacted in-person. The in-person contact included a vehicle inspection to provide visual verification of information collected during the interview, i.e., the presence or absence of damage to the car under study.

Respondent Universe

The survey universe consisted of all drivers of the approximately 98 million automobiles registered in the continental United States. The respondent was any eligible person in households contacted through telephone screening. The respondent need only have been the driver of an automobile owned or leased by the household for private use or owned by a company for business use.

The use of random digit telephone sampling excludes households without telephones from the sample. While this introduces some bias, it was not expected to be important. Unlike sampling from phone books, the use of a random digit dialing technique insures that households having unlisted telephone numbers have a chance of being selected. Households without telephones are more likely to have lower incomes and, therefore, more likely to have older vehicles that may be repaired less often. However, the more important fact is that bumper effectiveness is not different between households having and not having telephones.

Selection of Households

The households were selected using a cluster method of random digit dialing.[1] Clusters of 100 telephone numbers were selected by drawing a sample of eight digit numbers. Each number contained an area code/prefix combinations and two additional digits. Thus, this eight-digit number identified a potential 100 numbers formed from the selected eight digits and the two remaining numbers.

To the selected eight digits, a randomly selected number between 00 and 99 was appended. This number was called. If it was a residence, additional calls were made within the cluster until a predetermined number of residences were contacted. If the number was a business or nonworking number, the cluster was discarded and another selected. Initially, a cluster of 15 residences was sought. However, early in the surveying operation it became obvious that the rate of occurrence of low-speed incidents was less than anticipated. Two alternatives existed: add additional clusters; or, increase the cluster size. One of the advantages of telephone surveys over field surveys is the ease of adjusting the sampling rates to increase the yield. The cluster size was increased to 18 to obtain the desired sample size.

The cluster approach takes advantage of the telephone company's procedure of activating blocks of phone numbers. Simple random selection of numbers requires about five attempts to uncover one working residential number. Using the cluster approach, about three residences are uncovered for every five attempts. The sampling process selects telephone clusters in proportion to the (unknown) number of residences in the cluster. That is, a cluster with many residences is more likely to be chosen than one with few residences. The reduction in total calls needed is brought about because the phone company activates new numbers in banks of 100. Thus, there are many clusters for which no residential numbers have been assigned. These clusters are screened out early, resulting in a reduced workload.

2. DATA COLLECTION

In this section we review the results of attempted contacts. The initial and followup results are discussed separately.

Initial Contact

Table 1 presents summary statistics on response rates for the initial and followup contacts. In the process of identifying 10,326 residential numbers, 16,483 telephone numbers were dialed. The plurality of ineligible numbers were nonworking numbers. Of all the "No answer after five attempts," we estimate that approximately 85 percent were residential. These cases, although replaced with new numbers, have been used in the calculation of the initial contact nonresponse rate.

Of the 10,326 residential numbers contacted, approximately 28 percent of the households completed screenings and were eligible for the followup contact, i.e., the household had at least one licensed driver and the use of a car during the six-month recall period. Added to these households were 1,521 households (approximately 15%) which completed the screening but were ineligible for the followup contact. Of the 6,996 eligible households which completed the screening questionnaire, approximately 10 percent reported an eligible incident within the six-month recall period. These households were administered a main questionnaire. Of the incidents eligible for an extended interview, 99 percent resulted in completed main questionnaires.

The response rates for the initial and followup interview efforts are defined as follows: Screening response rate equals the number of screenings completed or partially completed with eligible and ineligible households divided by the number of residential households identified The main questionnaire response rate equals the number of main questionnaires completed divided by the number of eligible incidents reported.

Followup Contact

Table 1 also presents the summary statistics for the followup contact. Every eligible household that completed or partially completed an initial screening (4,949), excluding the households contacted during the initial supplement, was recontacted. Of those telephone numbers recontacted, 4,833 (nearly 98%) were still assigned to households. Of the 4,833 households contacted, approximately 93 percent of the households completed or partially completed a screening questionnaire. Only 37 households became ineligible at the time of the followup.

Of the eligible households which completed the screening questionnaire, approximately three percent reported an eligible incident within the followup period. These households were administered the main questionnaire. Aproximately 97 percent of the incidents eligible for an extended interview resulted in completed main questionnaires.

Table 2 distributes in finer detail the response rate of the followup effort. The completion rates for both screening and main questionnaires are appreciably higher for the telephone interviews than for the in-person interviews. Generally, in-person interviews meet with greater cooperation than telephone interviews. In this study, respondents were recontacted by telephone in order to schedule the personal followup visit. In effect, this provided the respondent with another opportunity to judge the interview as too burdensome and to refuse.

Table 1. Summary of response statistics

	Initial contact and supplement		Followup contact	
	Count	% of total	Count	% of total
Total telephone numbers attempted	16,483	100.0	4,949	100.0
Nonworking numbers	4,228	25.7	116	2.3
Nonresidential numbers	1,835	11.1	0	0.0
Residence out-of-county	11	.1	0	0.0
No answer after five attempts	548[a/]	3.3	0	0.0
Residential numbers dialed	10,326	62.6	4,833	97.7
Total residential numbers dialed	10,326	100.0	4,833	100.0
Completed screenings with eligible households	6,996	67.8	4,371	90.4
Partially completed screenings with eligible households[b/]	242	2.3	64	1.3
Ineligible households	1,521	14.7	37[c/]	.8
Screening response rate:		84.8		92.5
Refusals/breakoffs	860	8.3	177	3.7
Language problems	10	.1	0	0.0
No answer/nonresponse	465	4.5	60	1.2
Location problems	0	0.0	74	1.5
Other nonresponse[d/]	232	2.2	50	1.0
Screening refusal rate:		15.2		7.5
Total number of eligible incidents from completed/partially completed screenings	776	100.0	146	100.0
Main questionnaire response rate (completed questionnaires)	768	99.0	142	97.3
Other nonresponse	8	1.0	4	2.7

a/ We have estimated that approximately 465 (85%) of the 548 numbers called, where no answer was obtained after five attempts, were residential numbers. These 465 are included in the figure for residential numbers dialed as well.

b/ Partially completed households are those where screening information was gathered for at least one car in the household, but for some reason, information on one or more cars used by the household could not be obtained.

c/ These households sold or had use of a car during the six-month recall, but not during the followup period.

d/ Other nonresponse contains all types of nonresponse not included in other categories (e.g., respondent unavailable during field period).

Table 2. Summary of followup contact

	Telephone interview		In-person interview	
	Count	% of total	Count	% of total
Total telephone numbers attempted	4,527		422	
Nonworking numbers	116	2.6	0	0.0
Nonresidential numbers	0	0.0	0	0.0
Residential numbers dialed	4,411	97.4	422	100.0
Total residential numbers dialed	4,411	100.0	422	100.0
Completed screenings with eligible households	4,058	92.0	313	74.2
Partially completed screenings with eligible households	59	1.3	5	1.2
Ineligible households	32	.7	5	1.2
Screening response rate:		94.1		76.5
Refusals/breakoffs	121	2.7	56	13.3
No answer after five attempts	53	1.2	7	1.7
Location problems	55	1.2	19	4.5
Other nonresponse	33	.7	17	4.0
Screening refusal rate:		5.9		23.5
Total number of eligible incidents from completed/partially completed screenings	120	100.0	26	100.0
Main questionnaire response rate (completed questionnaires)	119	99.2	23	88.5
Other nonresponse	1	.8	3	11.5
Total number of eligible vehicle inspections from completed/partially completed screenings	-	-	492	
Vehicle inspection response rate (completed inspections)	-	-	480	97.6
Other nonresponse	-	-	12	2.4

3. BIAS DUE TO MEMORY RECALL

Of particular concern to this study is the trend in incident rates by month, as this might indicate bias through failure to remember an incident. In the initial interview, respondents were questioned about events occurring in the preceeding six months. If respondents were prone to forgetting past incidents, we would expect the reported incident rates to decrease for the more distant months of the recall period. We would also expect the incident rates for the shorter followup period to be equal to or higher than the most recent months of the six-month recall period, both because the followup period was shorter (approximately two months) and because of the interactive effect of the initial interview as a "reminder."

Table 3 shows the incident rates observed for each month of the recall and followup periods for the survey. Figure 1 is a graphical representation of the same data. Each rate is simply the number of eligible incidents for the specific month divided by the number of automobiles within the study which were in use for that month. The table shows the rates were lowest for recall months 4, 5, and 6, and also for followup months 2, 3, and 4. The drop in the rates suggests that limitations on respondents' memory do affect the reporting. The confidence limits displayed in the bottom rows of the table indicate that the differences in reported rates are significant; however, the estimated rates are not independent in that they are based on the same households. Only the most recent followup month has as high an incident rate as months 1, 2, and 3 of the six-month recall period. It is important to keep in mind that these monthly rates do not take into account actual road exposure, i.e., miles travelled. If we are to give credence to respondents' estimates of road exposure, the overall incident rate per 10,000 miles travelled (adjusted for time in study) is three times as high for the followup period (3.58) than for the recall period (1.13).

Table 3. Incident rates by month of study period: recall and followup periods

	Month prior to interview					
	1	2	3	4	5	6
Number of cars:						
Recall	10,658	10,535	10,402	10,264	10,110	10,208
Followup	6,480	6,466	1,398	143	10	0
No. of incidents:						
Recall	187	133	151	91	95	82
Followup	82	52	6	1	0	0
Incident rate:						
Recall	.018	.013	.015	.009	.009	.008
Followup	.013	.008	.004	.007	0	-
95% confidence interval for incident rate:						
Lower bound:						
Recall	.015	.010	.012	.007	.008	.006
Followup	.010	.006	.001	0	0	-
Upper bound:						
Recall	.020	.015	.017	.011	.011	.010
Followup	.015	.010	.008	.021	0	-

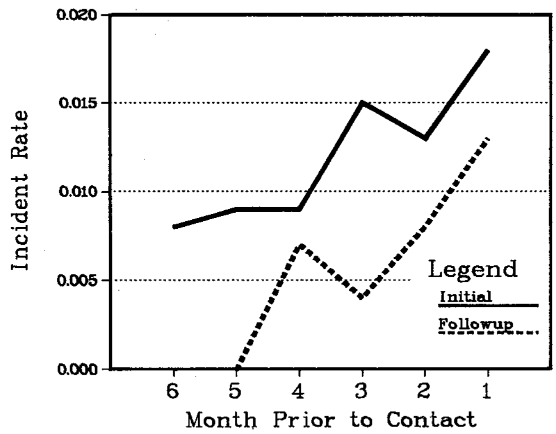

Figure 1. Monthly incident rates - recall and followup periods

Figure 2 graphically presents the same type of incident rates, displayed separately by model year class as well as for all cars combined for the recall period. The rates were separated in this manner to determine if a particular age group of vehicles showed reported rates that vary systematically over time. However, all model year classes of vehicles were found to have incident rates that were lowest for the most distant three months of the study period.

In the original plan, each household was to be contacted for re-interviewing two months after the case was first finalized. In practice, 10.7 percent of all vehicles were in study for more than two months during the followup phase. Figure 3 presents incident rates by model year class for the followup period. There is an across-the-board dropoff in incident rate over months one through three. Beyond the third month, the small sample sizes produce erratic behavior in the rates.

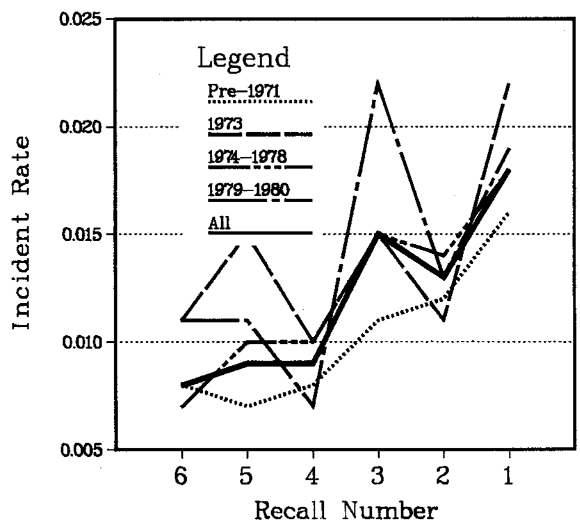

Figure 2. Monthly incident rates for cars of different model years - recall period

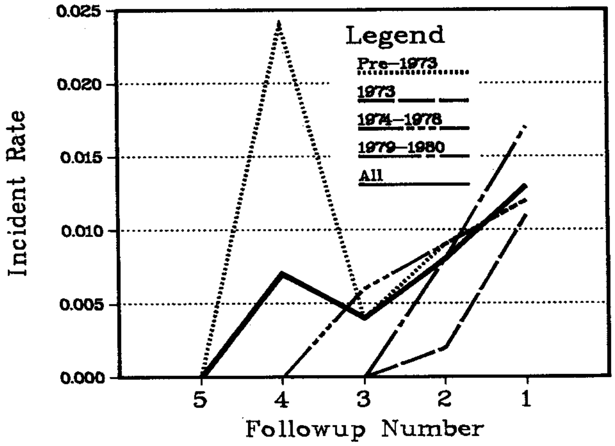

Figure 3. Monthly incident rates for cars of different model years - followup period

The Influence of Seasonality

The collection of supplemental screener records in the national study provided a unique opportunity to evaluate the assumptions about bias due to memory. Seasonal trends in the incidence of low-damage unreported accident involving bumpers may obscure the differences due to memory bias in estimates of incident rates for the recall and followup periods. With the use of supplemental screeners, there was a four-month period for which both recall and followup data were available. Incident rates computed from these two sources are directly comparable, without the complications of seasonality, and as such, their difference is a superior measure of bias due to memory.

Figure 4 is a graphical representation of incident rates and their associated 95 percent confidence bands for recall and followup data in the months of overlap. The distance between the two curves represents the bias. As can be seen, the distance is not constant over time, and the comparison must be judged inconclusive.

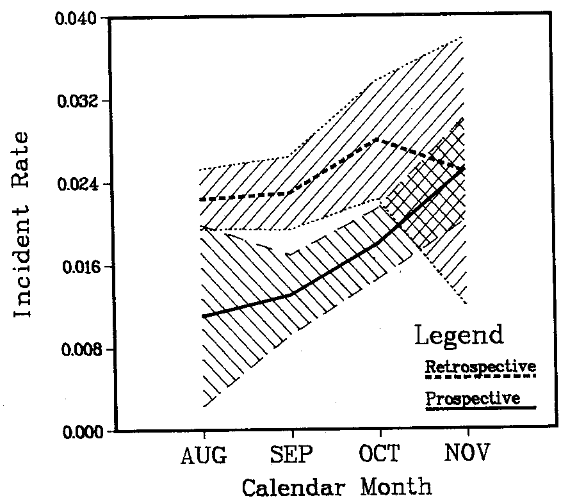

Figure 4. Ninety-five percent confidence bands for incident rates by calendar month - retrospective vs. prospective

Combining Bounded and Unbounded Interview Data

Each of the two phases of data collection, retrospective versus followup, has its strong and weak points as far as its usefulness in making estimates of incidence rates. The followup data will not have the bias of the retrospective data, but it will have a relatively higher variance because of the shorter collection period (two months). The retrospective data, while biased, will have a smaller variance because of a longer collection period (six months).

Let r represent an unbiased estimate of the incident rate as determined from the followup interviews and r' a biased estimate based on retrospective data. Then a linear combination,

$$r'' = wr + (1-w)r' \qquad 0<w<1$$

may be an estimate with less mean square error (MSE) than r alone if the proper w is selected. Denoting B as the bias of r', then

$$MSE(r'') = Var\ wr + (1-w)r' + (1-w)^2 b^2.$$

This is to minimize for

$$w = \frac{K + (1-\rho\sqrt{t})/t}{K + (t+1-2\rho\sqrt{t})/t}$$

where $K = B^2/Var(r)$ and t is the ratio of length of retrospective to prospective study. The average time in study for vehicles during the recall period was 5.47 months. For the followup period it was 2.11 months. Their ratio leads to a value of 2.59 for t.

To form the reduced MSE linear combination of estimates, we need estimates of the bias, B, of recall data and the correlation, ρ, of the followup and recall data. The estimated bias of the recall data may be taken as the estimated followup rate minus estimated recall rate. Assuming that the occurrence of the type of low probability event we are studying can be suitably modelled as a Poisson process, Var(r) may be estimated as equal to the incident rate r. The correlation ρ may be estimated as the sample correlation between two indicator variables I_r and I_f defined as follows: I_r equals one for vehicles with at least one eligible incident in the recall period; zero otherwise. I_f equals one for vehicles with at least one eligible incident in the recall period; zero otherwise. Based on our sample data, ρ is estimated as .02408.

Table 4 presents incident rates for the recall and followup periods and for the two periods combined.

Table 4. Rates for damage and no-damage incidents: recall, followup, and combined rates

Type of incident	Recall rate	Followup rate	B	Var(r)	W	Combined rate
With some damage reported	.005	.004	-.001	.004	.2738	.0047
With no damage reported	.007	.006	-.001	.006	.2738	.0067
All eligible	.012	.010	-.002	.010	.2739	.0114

USE OF LOG-LINEAR MODELS

Several variables were analyzed to determine the effectiveness of bumpers in preventing damage or reducing its extent. One of these, the probability of damage, is suitable for analysis using log-linear models. Four model year classes define the independent variable. A wide range of variables were considered as control variables and examined for possible masking effects. Below, we summarize this analysis.

The assumptions behind the distribution of test statistics used in log-linear analyses include a requirement for simple random sampling. Survey data rarely meet that assumption. In this particular survey, however, it is not an unreasonable assumption. Although households were selected via telephone clusters of 18 numbers, only those households reporting an eligible incident were included in the analysis. Approximately 530 clusters were selected and 909 incidents reported. Thus, the average cluster size was less than two. The effect on sampling errors $[1 + \rho(\bar{n}-1)]$ is quite small since the intraclass correlation (ρ) is expected to be near zero. (Estimated values of ρ are being computed and will be presented later.)

Bumper Effectiveness

The simplest indication that changes in bumper design have been effective is the declining damage rates observed across the model year classes, shown in Table 5 below. The 95 percent confidence intervals suggest that the pre-1973 cars (those built before any standards existed) experienced a significantly higher damage rate than either the 1974-78 or the 1979-80 classes. The overall rate, however, was not significantly different from the relatively small, 1973 class. The aggregate data suggest that damage rates are less for cars covered by any design requirements. However, the table ignores possible differences in the observed car classes. Below, we use log linear models to analyze various control variables, such as speed at impact, to determine if any important differences exist between model year classes.

Table 5. Probability of damage by model year class

Model year	Number of reported incidents	Number of incidents reporting damage	Percent of incidents with damage: Percent	St. Dev.	95% confidence interval: Lower	Upper
Pre-1973	281	150	53.4	3.0	47.6	59.2
1973	88	35	39.8	5.2	29.6	50.0
1974-78	449	164	36.5	2.3	32.0	41.0
1979-80	91	21	23.1	4.4	14.4	31.8
Total	909	370	40.7	1.6	37.5	43.9

The Log-linear Model

The data analyzed consisted of frequency counts arranged in a contingency table, where one dimension of the table, damaged versus not damaged, is viewed as a dependent variable. To perform the analytical computations, a statistical package, the Statistical Analysis System (SAS), was used. The program FUNCAT, based upon the work of Grizzle, Starmer, and Koch[2] was used to obtain the required values.

Each of the tables analyzed involved "model year" as an independent variable and one or two additional controls. Ideally, several variables might be considered simultaneously. Such a procedure can be analyzed easily by the SAS package. We have not shown more complicated tables (e.g., percent damaged by model year class by area of contact by traffic density, etc.) for two reasons. First, the number of possibilities is quite large. Second, the available data are too sparsely distributed to permit this. One restriction on the method is that all cells have positive probabilities. If the data are spread so thinly that cells are empty, the probabilities for these cells cannot be estimated confidently.

To further elaborate on the analysis, we will discuss one example where we define the frequency counts N_{ijkl} for a four-way table. The i indicates the degree of damage; j the model year class; and k and l are two control variables, say vehicle body type and location of accident. For this example, the various levels of each index will be as follows:

i = 1 damage
  = 2 no damage
j = 1 pre-1973
  = 2 1973
  = 3 1974-78
  = 4 1979-80
k = 1 subcompact
  = 2 compact
  = 3 intermediate
  = 4 full size
l = 1 roadway
  = 2 parking lot
  = 3 elsewhere

Thus, $N_{.jkl}$ is a count of all cars, damaged or not, in one model year class, body-type class, and accident location class. Therefore, $N_{1111}/N_{.111}$ is the proportion of vehicles damaged from the pre-1973 subcompact class where the incident occurred on the roadway. The log of the frequency counts is expressed as a linear model of the independent, control variables. For our example,

$$\log (N_{ijkl}) = u+a_i+b_j+c_k+d_l+ \text{interaction terms},$$

where a_i is a factor representing damage, b_j represents model year class, and so forth.

The first thing examined was the likelihood of an interaction between the a_i and b_j factors and the other control variables. If none existed, the analysis of damage versus model year class depended simply on the significance of the b_j. factor. If this factor was essentially zero, then the degree of damage did not vary with model year class. If, however, there exist significant interactions, the several combinations of factors must be examined individually to determine if the b_j factor is important conditionally, depending on values of the control variables.

An estimate of the magnitude of model year class differences is obtained from the value of b_j, and a confidence interval for this effect can be constructed from the SAS output.

Analysis of Damage Rates

In the analysis, the percent of cars damaged is examined across model year classes. Independence of the relationship between model year and bumper effectiveness with the control variable is indicated when the pattern of damage across model year groups is the same within each level of the control variable(s).

Independence from the control variable is important in this study for several reasons. For one, it permits less complicated conclusions to be drawn. The most important reason, however, deals with the question of speed at impact since the standards being evaluated only require protection in low-speed contacts. One of the most difficult aspects of the evaluation is to insure that attention is focused on this class of incident. To do this the survey instrument measured several variables related to the speed at impact. These include the driver's estimate of speed, as well as accident characteristics such as a description of the scene. If conclusions regarding bumper effectiveness are true when controlling for these "speed surrogates," then the conclusions are justifiable. If all of the speed-related factors are determined to be independent of the relation between percent damaged and model year class, then a clear conclusion can be reached about bumper effectiveness. In this manner, the confounding effect of the control variables can be removed.

The log-linear analysis led to the conclusion that all of the 13 control factors examined were independent of the relation between model year and percent damaged. While the initial computer runs included an interaction term between each of the two main effects (model year and one other control variable), the results indicated that the interactions were of little importance and were not needed. Thus, the log-linear model is a simple additive one, suggesting that the probability of damage can be obtained by multiplying an overall average by a factor indicating the control variable effect. The estimated models then contain only a main effect for model year and a main effect for the control variable. Exhibit 1 below contains a summary of the important computations. This exhibit contains the following:

- Col. (b) - the significance level of the control variable.
- Col. (c) - the significance level of the model year class variable.
- Cols.(d),(f),(h),(j) - the estimated coefficient for each model year class effect.
- Cols.(e),(g),(i),(k) the level of significance of each estimated coefficient.

Examining Col. (c) reveals that model year was a significant effect in every instance. Column (b) reveals that some of the control variables effect the probability of damage while others appear to have little impact. Variables which appear quite significant are: accelerating versus braking; striking versus struck; contact point bumper mismatch; relative motion of vehicles; location of accident, and estimated speed at impact. Variables with a possible effect are: front versus rear and vehicle size class.

[1] See Waksberg, Joseph, "Sampling Methods for Random Digit Dialing", Journal of the American Statistical Association, Vol. 73, No. 36, March 1978, pp 40-46.

[2] Grizzle, J. E., C. F. Starmer, and G. C. Koch, "Analysis of Categorical Data by Linear Models," Biometrics 25 (1969).

Control variable		Significance of model year categories	Model year class							
			Pre-73		73		74-78		79-80	
Name (a)	Level of significance (b)	(c)	Coefficient (d)	Level of significance (e)	Coefficient (f)	Level of significance (g)	Coefficient (h)	Level of significance (i)	Coefficient (j)	Level of significance (k)
Front/rear	.0398	.0001	.7513	.0001	.1887	.3334	.0475	.7120	-.9875	.0001
Accelerating/braking	.0063	.0274	.4369	.0763	.4913	.1887	-.3505	.1269	-.5777	.1157
Striking/struck	.0001	.0001	.6402	.0001	.6456	.7228	-.0526	.6526	-1.2332	.0011
Traffic density	.4819	.0001	1.0250	.0001	-.4888	.1441	-.0106	.9589	-.5256	.1741
Contact point	.0001	.0001	.6280	.0001	.1260	.5257	.0027	.9834	-.7567	.0014
Bumper mismatch by area of contact	.3260(area) .0018(mismatch)	.0001	.7205	.0001	.1925	.3615	-.0620	.6557	-.8510	.0007
Location of accident	.0078	.0001	.6359	.0001	.1168	.5412	-.0379	.7470	-.7148	.0004
Relative motion	.0037	.0001	.8107	.0001	.0697	.7726	.1556	.3403	-1.0360	.0014
Relative position	.4715	.0003	.7957	.0001	-.0272	.9194	.1151	.5108	-.6534	.0121
Vehicle size class	.0833	.0001	.7100	.0001	.1453	.4294	-.0559	.6374	-.7994	.0001
Type of object contacted	.3644	.0001	.7060	.0001	.1652	.4176	.0379	.7710	-.9091	.0002
Type of accident scene	.8411	.0001	1.1384	.0001	-.2288	.5660	.1015	.6533	-1.0111	.0226
Estimated speed	.0015	.0005	.8090	.0001	.0999	.7152	.1630	.3883	-1.0719	-.0056

Exhibit 1. Summary of log-linear analyses of percent damaged