# ALCOHOL AND TRAFFIC SAFETY: A SENSITIVITY ANALYSIS OF DATA FROM COMPOSITE SOURCES

M. A. CONNOLLY and A. W. KIMBALL
Johns Hopkins University, Department of Biostatistics, Baltimore, MD 21205, U.S.A.

and

L. H. MOULTON
University of Michigan, Department of Biostatistics, Ann Arbor, MI 48109, U.S.A.

**Abstract**—Risk factors associated with single-vehicle driver fatalities are explored in a sensitivity analysis of data from composite sources. Information on fatalities was taken from the Federal Accident Reporting System data base for 1976–1981. Characteristics of the driving population were given by the 1973 National Roadside Breath Testing Survey (Wolfe 1974). Using Bayes theorem and logistic regression analysis, the effect of changing driver characteristics on the probability of a fatality was explored. The method used is proposed for a case-control study in which the controls may not accurately represent the population from which the cases were drawn. Risk factors identified are generally in agreement with previous reports.

## 1. INTRODUCTION

Studies of the relationship between alcohol consumption and highway safety have spanned many years (Borkenstein 1985). Much knowledge has been gained, especially in the recent past, as a result of increased public awareness of the problem and a concomitant increase in funds, both public and private, to support research in the field. However, in one area, the epidemiology of traffic crashes, there are difficulties in design and execution (Simpson 1985). Determination of risk factors for traffic crashes is complicated by the lack of good information about the population at risk of a crash at the time and place a crash occurs. Case-control studies are complicated by the difficulties inherent in identification of suitable controls. Longitudinal studies are expensive and time-consuming because of the very low probability of an event for any individual at risk. Thus, it may be many years before substantial data from well-defined epidemiological studies become available.

Data sets are available now, however, that provide (1) multivariate information ($\mathbf{X}$) on cases (e.g. fatalities resulting from traffic crashes) and (2) multivariate information ($\mathbf{X}$) on drivers and vehicles not involved in crashes. From data set (1) one can estimate the probability of $\mathbf{X}$ given a fatality, and from data set (2) one can estimate the probability of $\mathbf{X}$ for the general driving public. Then knowing the overall probability of a fatality one can use Bayes theorem to calculate the probability of a fatality given $\mathbf{X}$. Data set (2) will not generally represent the precise population from which data set (1) was derived but it does provide a baseline from which variations can be taken.

This paper discusses the results of an analysis of the effects of variations in $\mathbf{X}$ on the overall probability of a fatality. Its major thrust is similar to what some investigators have called a sensitivity analysis, and it consists of two phases. In the first phase we determine the effect on the overall probability of a fatality of variations in $\mathbf{X}$ within limits dictated in part by the structure of data set (2). The second phase seeks to determine what range in variations in $\mathbf{X}$ would yield the fatality rate actually observed for the period and location under study. That is, it provides ranges of combinations of the several variables, any of which might have yielded the number of fatalities observed. Thus, the population at risk, although not known, is circumscribed.

Sections 2 through 5 concern steps to define and estimate parameters of the sensitivity analysis. In Section 2 we discuss data sources and sample selection. In Section 3 we list the variables that are common to the two data sources, for drivers and fatalities,

and present tables and descriptive statistics. Results of logistic regression analysis used to identify risk factors among these variables are in Section 4. In Section 5 we define the variables and categories of the sensitivity analysis and outline steps to estimate cell conditional fatality rates that are parameters of the sensitivity analysis. The inverse probability (Bayes) equation is required in this step.

Sections 6 through 9 describe the sensitivity analysis. Objectives, and numerical methods to change one or two marginal distributions while maintaining observed structure and cell conditional fatality rates, are covered in Section 6. In Section 7 we describe the types of outcome to be measured. These are changes to the average fatality rate as a single margin is varied and changes to a second margin that will restore the observed average fatality rate. Tables and results of the sensitivity analysis are presented in Section 8. In Section 9 we summarize and discuss implications of these results. Section 10 is a discussion of the methods and their limitations.

## 2. DATA SOURCES AND SAMPLE SELECTION

Although relations between alcohol use and all aspects of highway safety are of interest, the characteristics of drivers and passengers involved in fatal accidents are better documented than are characteristics of any other drinking and driving population. The Fatal Accident Reporting System (FARS) data base contains reports of all fatalities involving a motor vehicle and occurring on a public road in the United States from 1975 on. Records from 1976 to 1981 were used to characterize the fatalities in this study. For each person involved in a fatal accident are recorded age, sex, blood alcohol (BAL), seat belt use, and, if the person was a fatality, the date of death and interval from accident to death. The time, place, road type, and conditions of each accident are recorded. In addition, characteristics of the vehicle(s), year, weight, and defects related to the accident are available.

Description of the driving population might come from several sources, none complete. The Department of Transportation (DOT) publishes Highway Statistics yearly, which concern highway use, vehicle miles of travel, and minimal driver description (sex, age) all by state and month. The DOT also sponsored a 1977 Nationwide Personal Transportation Survey and published the reports: *Vehicle Occupany* (Kusmyak 1981a), *Household Travel* (Klinger et al. 1982), and *Household Vehicle Utilization* (Kusmyak 1981b). Such statistics may provide gross denominator data for the fatality rates but do not concern alcohol and are not easily comparable to the FARS data. A few surveys of drivers have been done. The two largest were a community survey by Borkenstein reported in 1964 and the preliminary survey, for an intended nationwide study, conducted by A. C. Wolfe in 1973. This second survey, the National Roadside Breathtesting Survey (NRBS) measured driver and trip characteristics for 3,698 interviews and breath tests at 185 sites in 18 states. All cars were stopped between 10 P.M. and 4 A.M. (most before 3 A.M.) on a Friday or Saturday night in the eight weekends of November and December in 1973. These data were judged highly relevant to the intent of this study and were made available to us by the author. Interviewers for the NRB Survey recorded driver's age, sex, trip reason and length, blood alcohol (breath test), vehicle type, seat belt use, and site characteristics. Components of the last item, except for weather, were determined by the survey design.

Baseline values in the sensitivity model were derived from unweighted frequencies in comparable subsets of FARS and NRBS data, from the 1981 DOT publications mentioned in the previous paragraph, and DOT summaries of licensed drivers by year and state. DOT statistics were used to estimate fatality rates in the general population but these did not affect subsequent estimates of relative fatality rates for different groups. For comparison to the information available in FARS, DOT statistics were also used to estimate the proportion of cars in the NRBS population that were small, mid-sized, or large.

The risk set described by the NRB Survey comprised trips taken during weekend nights in the late fall/early winter. The relevant events in the FARS sample were fatal

accidents that occurred on weekend nights to occupants of noncommercial vehicles. Preliminary analysis showed that the age, sex, and vehicle occupancy distribution of the FARS sample did not vary significantly by month and so the selection of FARS records included all seasons. Since driver blood alcohol and responsibility were of primary interest it was decided to select only single-vehicle driver fatalities. Single-vehicle accidents are more directly attributable to error by the single driver, and cases in which a passenger dies but the driver survives may not report driver BAL as accurately as is the case for driver fatalities. Also, such accidents require passengers, and so might tend to average more occupants than is representative of the population at risk. Single-vehicle driver fatalities are not known to be correlated with the number of passengers in the car or with passenger fatalities.

Alcohol reporting for FARS is variable across states. In preliminary analysis of FARS data, the 15 states with the highest reporting rates were chosen from the 52 states and territories in FARS. For the sensitivity analysis model, 6 states from the 15 in common with the 18 states surveyed by NRBS were selected. Table 1 lists these states, their licensed populations, and fatality counts for 1982, showing that they represent approximately one-fifth of the total U.S. population, licensed drivers, and fatalities. There were 13,704 single-vehicle nonstationary accidents in the FARS data that occurred in the six states in 1976 through 1981. This number does not include any accidents involving motorcycles or unknown vehicle types. Of these cases, 10,266 were driver fatalities. When cases involving buses or tractor trailors were excluded, the number of accidents in the sample that occurred between 10 P.M. and 3 A.M. on Friday or Saturday nights was 1,838. All of these were complete for driver age and 1,555 reported driver BAL.

The NRBS data had a total of 3,698 interviews. Of these, 970 were in the six states that had high BAL reporting rates in FARS. When records missing vehicle body type, driver age, or sex were excluded, the NRBS sample size was 944. Of these, 897 cases were complete for BAL.

## 3. VARIABLE SELECTION

Variables to be compared between the populations were driver's age, sex, and BAL, number of occupants, vehicle body type, weather, day, and time. These were the only

Table 1. Population and number of fatalities by state

| State | Population* (millions) | Number licensed drivers** (millions) | Number of*** Fatalities |
|---|---|---|---|
| California | 23.67 | 16.3 | 4,618 |
| New Jersey | 7.40 | 5.34 | 1,086 |
| Oregon | 2.63 | 1.89 | 518 |
| Virginia | 5.35 | 3.63 | 879 |
| Washington | 4.13 | 2.77 | 748 |
| Wisconsin | 4.71 | 3.04 | 770 |
| % of U.S. total | 21 | 22 | 20 |

* - from 1980 census, total U.S. population = 226.5 million

** Highway statistics, 1982 by U.S. Department of Transportation, Federal Highway Administration, No. 38 in series

*** Highway statistics, 1982 - total in U.S. for that year: 43,862

Table 2a. Age of driver by population

|  | N | Mean age(yrs.) | Median | Std. dev. | Range |
|---|---|---|---|---|---|
| FARS | 1,838 | 27.6 | 23 | 11.6 | 14-84 |
| NRBS | 944 | 30.5 | 26 | 13.0 | 16-79 |

Table 2b. Distribution of males by population

|  | % | N |
|---|---|---|
| FARS | 85.4 | 1,838 |
| NBRS | 81.1 | 944 |

Table 2c. BAL of driver by population

|  | N | Mean(100 gm/ml) | Median | Std. dev. | Range |
|---|---|---|---|---|---|
| FARS | 1,555 | .16 | .17 | .858 | 0-.63 |
| NRBS | 897 | .018 | 0 | .35 | 0-.26 |

Table 2d. Seat belt use by population

|  | % missing | N |
|---|---|---|
| FARS 2.0 % using | 44.0 | 1838 |
| NRBS 27.1 % not using | 10.1 | 944 |

Table 2e. Percent distribution of vehicle occupancy by population

|  | Number of occupants | | | | |
|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 or more | N |
| FARS | 59.5 | 26.0 | 8.4 | 6.0 | 1,838 |
| NRBS | 34.9 | 36.0 | 13.8 | 15.4 | 944 |

Table 2f. Percent distribution of vehicle body type by population

|  | Percent in each category | | | |
|---|---|---|---|---|
|  | Car | Cargo | Recreational | Number |
| FARS | 75.9 | 20.9 | 3.2 | 1,838 |
| NRBS | 88.8 | 8.7 | 2.5 | 944 |

Table 2g. Distribution of dry weather conditions by population

|      | %    | N     |
|------|------|-------|
| FARS | 88.6 | 1,838 |
| NRBS | 76.4 | 944   |

Table 2h. Distribution of Friday night trips by population

|      | %    | N     |
|------|------|-------|
| FARS | 52.4 | 1,838 |
| NRBS | 47.8 | 944   |

Table 2i. Percent distribution of time of night by population

| | Time of Night | | | | | |
|------|-------------|----------|-------------|--------|--------|-------|
|      | 10-10:59 pm | 11-11:59 | 12-12:59 am | 1-1:59 | 1-2:59 | N     |
| FARS | 13.8        | 16.1     | 18.6        | 24.2   | 27.3   | 1,838 |
| NRBS | 25.3        | 26.3     | 1.5         | 15.1   | 21.9   | 944   |

Table 2j. Percent distribution of number of lanes by population

|      | 1 lane | 2 lanes | 3+ lanes | Unknown | N     |
|------|--------|---------|----------|---------|-------|
| FARS | 0      | 36.1    | 63.8     | 0       | 1,838 |
| NRBS | 1.25   | 82.0    | 15.7     | 1.03    | 944   |

Table 2k. Percent distribution of subjects by state and population

|      | CA   | NJ   | OR    | VA   | WA   | WI   | N     |
|------|------|------|-------|------|------|------|-------|
| FARS | 13.4 | 43.6 | 13.2  | 5.9  | 14.3 | 9.6  | 1,838 |
| NRBS | 15.7 | 13.2 | 14.2. | 10.7 | 32.6 | 13.7 | 944   |

Table 2l. Percent distribution of fatalities by year

|      | 1973 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | N     |
|------|------|------|------|------|------|------|------|-------|
| FARS | 0    | 11.2 | 16.0 | 19.5 | 19.0 | 18.3 | 15.9 | 1,838 |

relevant variables common to the two data sets. Distributions with respect to the number of lanes and to the state where interviews or fatalities occurred were explored, but it was thought best to avoid using variables in the model that could be highly dependent on the selection process used by the NRBS. The road size, body type, and weather variables must be regarded with a degree of caution due to the arbitrary decisions made on how to code them in a uniform manner across the two data sets.

The FARS and NRBS samples were selected to have complete reporting for vehicle body type. The FARS sample was selected to exclude motorcycles, large trucks, and buses. The sample from NRBS happened not to have any motorcycles or large commercial vehicles. Although FARS records included information about vehicle curb weight, the NRBS coding allowed only three classes to be distinguished: cars, light trucks, and recreational vehicles.

The NRBS weather description distinguishes continuous from intermittent rain and/ or snow, and other categories are clear or foggy conditions. The FARS coding distinguishes types of precipitation but not intensity and has the categories clear and foggy. For comparison, these conditions were classed as wet (any precipitation) compared to dry. Logistic regression models were used to measure association between trip characteristics and inclusion in either the FARS or NRBS sample. These were prospective models, the outcome being status as a NRBS(0) or FARS(1) case. Covariates were those variables in Table 2 with similar coding in the two samples that were not associated with the NRBS design. Included in the model were continuous covariates for age and BAL and ordinal effects of number of occupants (1, 2, 3, and 4 or more). The reference group for seat belt used, not used, and missing was the not-used category. Weather was coded $0 = $ dry, $1 = $ wet. The reference class for vehicle type was automobile. Sex was coded $0 = $ male, $1 = $ female. The best set of predictors was chosen from models that included main effects and possible two-way interactions. A second logistic regression model had entirely categorical covariates and included only those variables used in the sensitivity analysis.

As shown in Table 2a, the age range of drivers in the NRBS sample (16 to 79 years) is almost as great as that of the FARS sample (14 to 84 years). The breathtested drivers average three years older than the fatalities. The proportion of males is similar between the samples (Table 2b). Blood alcohol data (Table 2c) is available for 84.6% of the selected FARS cases and for 95.0% of selected NRBS cases. The blood alcohol measured on fatalities averaged .16 g/dL. In the highway survey the average was .018 g/dL. The median blood alcohol for fatalities was .17 g/dL. For the NRBS sample, median blood alcohol was 0 overall, and for males in the sample the median was .01.

Seat belt use was a variable of interest (Table 2d). In the FARS set, the rate of use was 3.6% among cases that had a reported value, but 44% of the cases were missing the value for seat belt use. The missing rate does not change over time but, as shown in Table 3, the six states have reporting rates that range from 90% missing (California) to 1.2% missing (Virginia). Within the five better reporting states, seat belts were worn by 2.84% of the FARS sample among the 92% of cases reporting use. Belt use was included in the model despite the high rate of missing reports.

The number of occupants in a vehicle tends to be lower for the FARS sample (Table 2e). The majority (59%) of single-vehicle driver fatalities had only the driver present. This sample of single-vehicle driver fatalities has a different distribution with respect to number of occupants than does the larger set of single-vehicle fatalities. The latter has more vehicles with several occupants than does the NRBS sample.

Table 2f shows the distributions of vehicle body types. FARS coding included automobile curb weights but NRBS coding did not distinguish weight classes. The cargo class represents light trucks. Vans and other recreational vehicles are in the third class. Of the FARS sample, 20.9% were in trucks and 8.7% of the NRBS sample were in this class. Because of the small proportions in the third category, 3.2% of FARS and 2.5% of NRBS, this was combined with the second category for the sensitivity analysis.

Results in tables 2g–l are influenced by the NRB Survey design. Table 2g shows that dry conditions are more common in the FARS than in the NRBS sample. The NRB

Table 3. Seatbelt use reporting rates in FARS by state

| State | N | % reporting on belt use | Seat belt use | |
| | | | % of total | % of reporting |
|---|---|---|---|---|
| CA | 802 | 9.85 | 1.25 | 12.66 |
| NJ | 175 | 97.73 | 2.84 | 2.91 |
| OR | 108 | 93.52 | 1.85 | 1.98 |
| VA | 246 | 98.78 | 3.66 | 3.70 |
| WA | 243 | 88.06 | 2.46 | 2.80 |
| WI | 263 | 79.85 | 1.90 | 2.38 |
| Total | 1,838 | 55.98 | 2.01 | 3.60 |
| Total excluding CA | 1,036 | 91.70 | 2.61 | 2.84 |

Survey was during late fall while the FARS reports accidents at any time of year. All accidents and driver surveys were on Friday or Saturday nights. Table 2h shows that FARS had a slightly higher proportion of cases on Friday than did NRBS. The NRB Survey was designed to collect one sample between 10 P.M. and 12 midnight and another between 1 and 3 A.M. The sampling rate was based on expected traffic volume. If interviewers found that their sample was smaller than intended, then they could continue to sample slightly longer in either period. Table 2i shows few observations in the hour after midnight and none after 3 A.M. for this sample from the NRBS data. Coding for number of lanes was very different between FARS and NRBS. The FARS count referred to the number in either direction on one side of a dividing strip. The NRBS categories include count of lanes in one direction and an indicator for any median strip. No NRBS interviews were done on extremely small roads or on limited access highways. Table 2j shows these differences based on an abbreviated coding.

In Table 2k are the relative distributions in the six states for the two samples. Although the NRBS sample is proportional to regional populations, it may be seen by comparing Table 2k to Table 1 that Washington State is overrepresented in the NRBS sample. Table 2l shows the distribution over the years 1975 to 1981 of the FARS sample. The NRBS sample was collected in 1973.

## 4. RESULTS OF LOGISTIC REGRESSIONS

The logistic regression coefficient for a trip characteristic, such as a female driver, estimates the log of the odds ratio for the probability that an event in the FARS sample has a female driver compared to the probability that an NRBS trip has a female driver. If the NRBS sample represents the population at risk for the FARS sample, then the logistic regression coefficients also estimate the increased probability that a trip by a female driver will end in a fatality compared to the risk for a male driver. In any case, the logistic regression estimates reflect differences in the multivariate distributions of trip characteristics between the two samples. Driver BAL, seat belt use, age of driver, number of occupants, and vehicle body type had different distributions in the FARS and NRBS samples. Two other characteristics that distinguished the samples, number of lanes and distribution of subjects over the states, may reflect differences between the NRBS design and the FARS sample.

Coefficients and odds ratios estimated on logistic regressions with BAL, age, and number of occupants entered as continuous or ordered covariates are reported in Table 4. In the regression equation on the left side of the table were driver BAL, missing seat belt, number of occupants, driver age, belt use, cargo, and wet weather listed in decreasing order by significance. The high estimated odds ratio (5.81 compared to belt not used) for the belt use not reported category reflects a higher nonreporting rate in FARS. The odds ratio for belt used vs. not belt used is .23. Odds ratios calculated from logistic regression coefficients are 4.57 for each .05 g/dL increase in BAL, and .64 for wet compared to dry weather. The odds ratios for cargo and recreational vehicles compared to automobiles are 1.96 and 1.52. Lower risks are associated with increased age as the odds ratio is .80 for each 5-year difference (such as 20 vs. 25). Trips with more occupants also have lower risk. The odds ratio for 3 vs. 2 or 2 vs. 1 is .57. Among these effects, there is less evidence that odds ratios for recreational compared to passenger vehicles or the sex effect are in fact different from zero.

All of the logistic regression covariables of Table 4, except weather, were included in the sensitivity analysis model. Although it had a significant regression coefficient, weather was not included in the sensitivity analysis because the coding for this variable was very different between FARS and NRBS cases, and because its distribution in the NRBS sample was strongly influenced by the survey design. Sex was included in the sensitivity analysis because BAL distributions are quite different in male and female drivers. Also, it may be noted that although the proportion of males is higher in the FARS than in the NRBS sample, the logistic regression coefficients suggest higher risk for female drivers (Table 4). The six variables of the sensitivity analysis model are age,

Table 4. Odds ratios and 95% confidence intervals from the logistic regressions*, $N = 2,452$ (897 in NRBS, 1,555 in FARS)

| Covariate | Odds ratio | Confidence interval |
|---|---|---|
| BAL(for each increase of 0.05 gm/dℓ) | 4.57 | (3.98, 5.25) |
| Age (for each increase of 5 yrs.) | .80 | (.76, .86) |
| Number of occupants (for each increase of one) | .57 | (.49, .66) |
| Weather | .64 | (.44, .93) |
| Cargo | 1.96 | (1.29, 2.98) |
| Recreational | 1.52 | (.63, 3.68) |
| Belt use | .23 | (.14, .39) |
| Belt use not reported | 5.81 | (4.10, 8.24) |
| Sex | 1.41 | (.97, 2.05) |

*Codes for non-ordinal variables:

Weather (0 = dry, 1 = wet), vehicle type (reference group for cargo and recreational is automobile), seatbelt (reference group for belt used and belt not reported is seatbelt not used), sex (0 = male, 1 = female)

Alcohol and traffic safety

Table 5. Odds ratios and 95% confidence intervals from the logistic regression with all covariates categorized

| Include missing seat belt N = 2,452 (897 in NRBS) | | | Use cases complete for seat belt N = 1,648 (806 in NRBS) | | |
|---|---|---|---|---|---|
| Covariate | Odds ratio | Confidence Interval | Covariate | Odds ratio | Confidence Interval |
| BAL = 0 | 1.00 | – | BAL = 0 | 1.00 | – |
| 0<BAL<.02 | .09 | (.03-.23) | 0<BAL<.02 | 0.10 | (.03-.32) |
| .02<BAL<.05 | 1.60 | (.96-2.67) | .02<BAL<.05 | 1.51 | (.81-2.80) |
| .05<BAL<.10 | 6.82 | (4.60-10.11) | .05<BAL<.10 | 7.32 | (4.68-11.44) |
| .1<BAL<.15 | 31.2 | (19.3-50.4) | .1<BAL<.15 | 33.1 | (19.3-5.68) |
| .15<BAL<.2 | .78 | (83-380) | .15<BAL<.2 | 187 | (82-424) |
| .2<BAL | 478 | (177-1,289) | .2BAL | 455 | (169-1,224) |
| Age under 20 | 1.00 | – | Age under 20 | 1.00 | – |
| 20<Age<30 | .59 | (.41-.84) | 20<Age<30 | .66 | (.44-.97) |
| 30<Age<40 | .28 | (.17-.45) | 30<Age<40 | .21 | (.12-.37) |
| 40<Age | .18 | (.11-.29) | 40<Age | .19 | (.10-33) |
| Male | 1.00 | – | Male | 1.00 | – |
| Female | 1.34 | (.91-1.95) | Female | 1.22 | (.80-1.87) |
| 1 occupant | 1.00 | – | 1 occupant | 1.00 | – |
| 2 occupants | .50 | (.36-.70) | 2 occupants | .54 | (.37-.79) |
| 3+ | .27 | (.18-41) | 3+ occupants | .28 | (.18-45) |
| Automobile | 1.00 | – | Automobile | 1.00 | – |
| Non-auto | 1.99 | (1.33-3.00) | Non-auto | 1.65 | (1.01-2.69) |
| Seat belt not not used | 1.00 | – | Seat belt not not used | 1.00 | – |
| Belt used | .24 | (.14-.42) | Belt used | .23 | (.13-.40) |
| Belt use missing | 6.05 | (4.22-8.68) | | | |

sex, BAL, seatbelt use of the driver, number of occupants, and vehicle body type. Because the sensitivity analysis describes probabilities in a multidimensional table, all covariates were coded as categorical. Table 5 reports logistic regression estimates for all categories of the model.

The categorical coding of Table 5 has age at four levels (ages under 20, up to 30, up to 40, and over 40), number of occupants at three levels with the three and over categories combined, and BAL in seven categories. The regression on the left of Table 5 included 804 observations missing belt use and has an estimate of the coefficient for the missing category. Results on the right of Table 5 are based on the sample that had no missing seat belt use values. All BAL categories, except for the range of .02 to .05 mg/dL, appeared to have significantly different risks from BAL = 0. The apparent protective effect of the lowest nonzero BAL category may be partially due to coding difference between samples or to nonresponse bias. Because NRBS coding used units of .001 mg/dL and FARS coding was usually rounded to units of .01, the second BAL category represented only a single point among possible FARS values. Other coefficients of Table 5 indicate a protective effect of all older age categories compared to teen-aged drivers, higher risk for female drivers, for vehicles with one occupant, and for trucks and vans compared to automobiles. Seat belt uses had a lower risk than nonusers and the missing category had a higher risk than nonusers. (A higher proportion of FARS than NRBS cases were missing this value.)

Table 6. Categories for the sensitivity analysis

| Ages | Sex | Number of occupants | Vehicle body type | Seat belt use | Blood alcohol level (g/dL) |
|---|---|---|---|---|---|
| <20 | Male | 1 | Small car | Seat belt | Between 0 and .02 |
| 20–30 | Female | 2 | Medium-sized car | not used | .02–.05 |
| 30–40 | | 3+ | Large car | Seat belt | .05–.1 |
| 40+ | | | Other vehicles | used | .1–.15 |
| | | | | | .15–.2 |
| | | | | | .2–.25 |
| | | | | | .25–.30 |
| | | | | | .30–.35 |
| | | | | | .35+ |

## 5. ESTIMATION OF CELL CONDITIONAL FATALITY RATES

The sensitivity analysis is based on data for the logistic regression models of the previous section. The objective is to define ranges of distributions of the covariable set that are consistent with the data. The first step is to construct tables of probabilities for the different trip variables among the NRBS sample of trips and among the FARS sample of fatalities. Probabilities for classes of a single covariate (e.g. age or sex) may be estimated directly from observed proportions. However, the samples are not large enough to give nonzero estimates for all combined categories, such as the probability of a trip with a 30- 40- year-old female driver in a pick-up truck using a seatbelt with BAL between .05 and .1 g/dL with two passengers. A "smoothing" procedure was used to create tables of probabilities with the marginal categories of Table 6 that had positive probabilities in each of the 1,920 cells that are implied by the categories of Table 6.

Let $F$ = fatality and $X_{ijk}$, or $X$, label the characteristics of a particular trip (the six trip descriptors would properly be labeled by six subscripts). Dot notation will be used to denote marginal categories of a characteristic, e.g. $X_{i.}$ will represent the $i$th category of the first character. From the NRBS we obtained $Pr(X)$, the probability that a driver

Table 7. Observed frequencies by age, sex, number of occupants, and population

| NRBS (N = 944) | | | FARS (N = 1,836) | | |
|---|---|---|---|---|---|
| **1 occupant** | | | | | |
| **Age** | **Male** | **Female** | **Male** | **Female** | |
| under 20 | 62 | 13 | 172 | 33 | |
| 20–30 | 91 | 29 | 404 | 66 | |
| 30–40 | 45 | 14 | 178 | 32 | |
| 40 & over | 55 | 20 | 175 | 34 | |
| **2 occupants** | | | | | |
| under 20 | 54 | 17 | 134 | 26 | |
| 20–30 | 110 | 20 | 191 | 23 | |
| 30–40 | 43 | 7 | 50 | 14 | |
| 40 & over | 73 | 10 | 35 | 5 | |
| **3 or more occupants** | | | | | |
| under 20 | 47 | 9 | 89 | 16 | |
| 20–30 | 83 | 17 | 110 | 11 | |
| 30–40 | 53 | 11 | 22 | 3 | |
| 40 & over | 50 | 5 | 9 | 5 | |

at risk of a fatality has characteristics **X**. From the FARS data set we calculated $\Pr(\mathbf{X}|F)$, the probability that a driver has characteristics **X** given that a fatality has occurred.

The smoothing procedure to estimate positive values for $\Pr(\mathbf{X})$ and $\Pr(\mathbf{X}|F)$ was based on average and conditional probabilities. First, for both $\Pr(\mathbf{X})$ and $\Pr(\mathbf{X}|F)$ observed frequencies in 24 cells of the age by sex by number of passenger table were used to estimate corresponding marginal probabilities. NRBS data were used to estimate $\Pr(\mathbf{X})$, and FARS data to estimate $\Pr(\mathbf{X}|F)$. These frequencies, reported in Table 7, were based on all available observations, including those missing BAL or other information. Next, the distribution of BAL within each data set was examined in order to find an appropriate model. In the NRBS data, it appeared that the logarithm of the proportion in the tail of the BAL distribution above some blood alcohol level $\beta_0$ decreased linearly with $B$ (Fig. 1). Thus the tail of the distribution was modeled as $\Pr(\text{BAL} > B|B > \beta_0) = \exp(-\lambda B)$ for the survey drivers. There were sufficient observations in the first two BAL categories, $\Pr(\text{BAL} = 0)$, and $\Pr(0 < \text{BAL} < .02)$, of each cell in the three-way table to estimate these probabilities by their frequencies. The last eight probabilities for BAL conditional on age, sex, and number of passengers were fit by estimating $\lambda$ within each cell, fitting tail probabilities according to the exponential model, and normalizing so that conditional probabilities over the 10 BAL categories summed to one.

The BAL distribution among fatalities had a different shape from that among surveyed drivers (Fig. 2). The FARS distribution had positive probability at BAL = 0 and the distribution of nonzero values wre nearly symmetrical about their mean. The observed frequency of BAL = 0 at each category of the three-way table was taken as the estimate of that probability. The mean and standard deviation ($\mu_{ijk}$, $\sigma_{ijk}$) of the positive BAL values were estimated within each sex by age by number of occupants classification. The other nine BAL categories were fitted based on normal ($\mu_{ijk}$, $\sigma_{ijk}$) integrals between cut points. The nine probabilities for positive BAL were then normalized so as to sum to the estimate for $1-\Pr(\text{BAL} = 0|\text{sex, age, number of occupants})$.

For this comparison of FARS and NRBS distributions, vehicle body type classification includes three automobile size classes and a fourth class for trucks, recreational
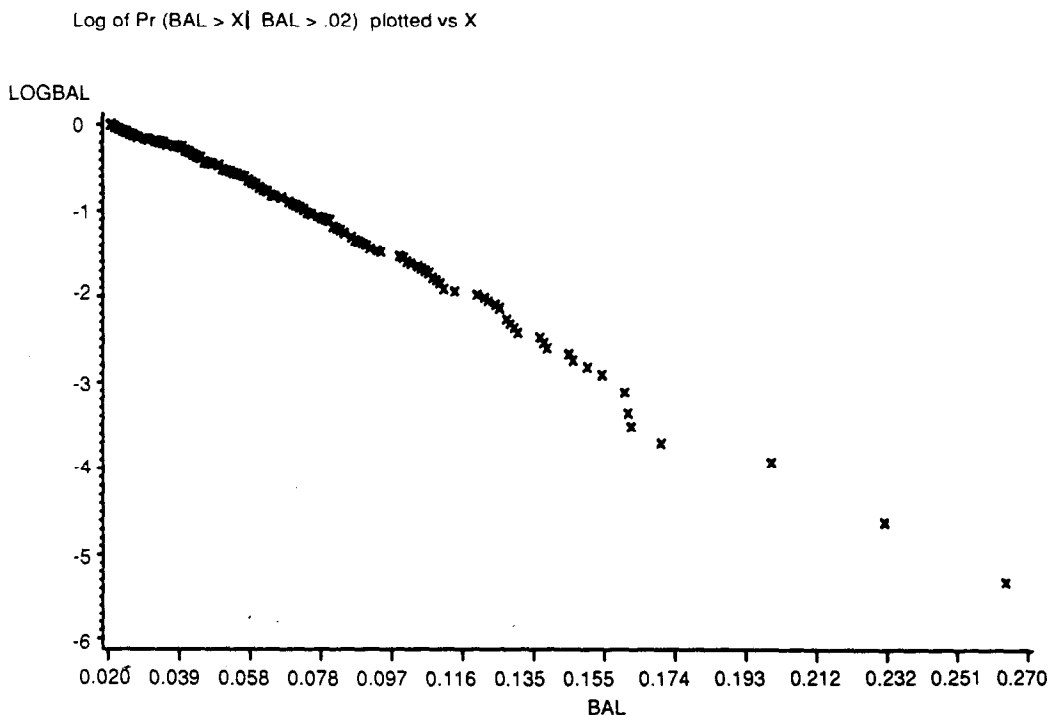
Log of Pr (BAL > X| BAL > .02) plotted vs X



Fig. 1. Logarithm of NRBS BAL tail probability for BAL >2.

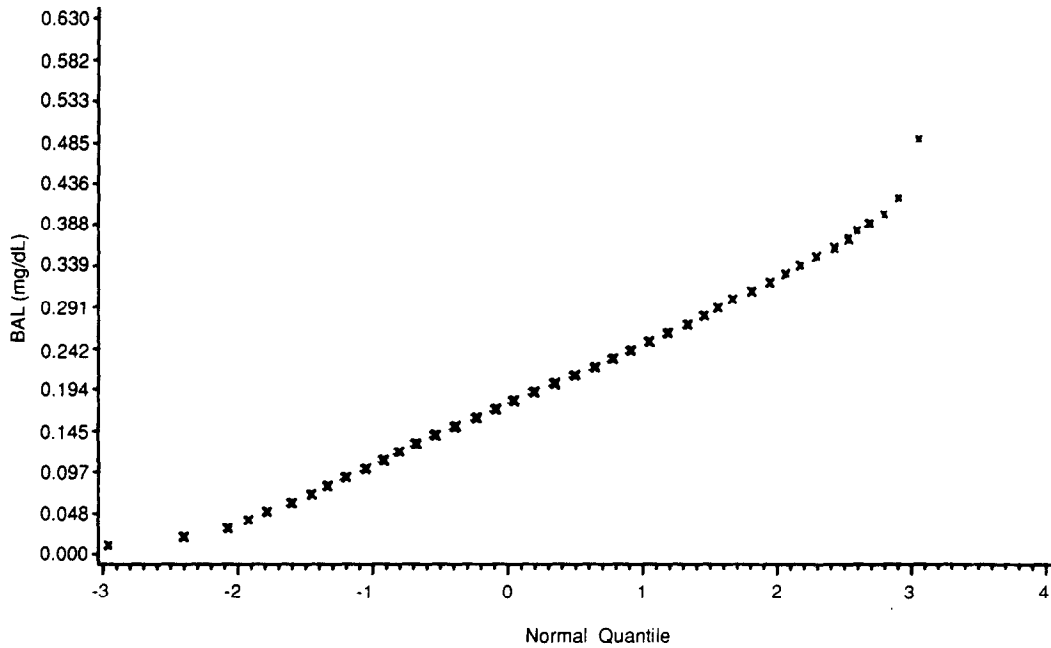BAL values above 0 vs quantiles of normal (0, 1)



Fig. 2. Quantile–quantile plot of FARS BAL distribution.

vehicles, or vans. The FARS data included information about automobile curb weight as well as distinctions among cars, light trucks, and vans. The NRBS data did not distinguish among automobile size classes, so the DOT report *Household Vehicle Utilization* (Kusmyak 1981a) was used to augment the NRBS data for comparison with FARS vehicle classification. The classes were small cars (2,500 lb.), midsize (2,500 to 4,000 lb.) and large cars (4,000 lbs. and over). The percentages of all miles driven by each of these classes, based on the DOT report, were 20.5%, 39.5%, and 40%, respectively. Observed proportions of drivers of trucks or recreational vehicles in each age by sex class of the NRBS data were used to estimate probabilities in the fourth vehicle body type classification, and drivers were assumed to be distributed in the three auto size classes according to the above listed proportions. FARS probabilities in the four vehicle classes were fitted according to frequencies observed in the age by sex classification.

Seat belt use appeared to vary with driver's age but to be independent of driver's sex or the number of occupants. The average belt use rate in each age grouping was used to estimate cell probabilities. These conditioning steps imply that any associations of vehicle type, driver BAL, and seat belt use are through their conditional distribution given age by sex by number of occupants classification.

These smoothing procedures were applied to the FARS and NRBS samples to create corresponding tables of $\Pr(X_{ijk}|F)$ and $\Pr(X_{ijk})$. The Bayes equation to estimate $\Pr(F|X_{ijk})$, the probability of a fatality for a given set of trip characteristics is

$$\Pr(F|X_{ijk}) = \Pr(X_{ijk}|F) \cdot \Pr(F)/\Pr(X_{ijk}). \tag{1}$$

The value of $\Pr(F)$ in eqn (1) was estimated from United States Department of Transportation sources. The total number of licensed drivers in the six states was reported in DOT summaries of 1976–1981. The average number of trips per driver per week in the hours of this study was estimated to be .222 based on statistics in the 1977 National Personal Transportation survey publications, *Household Travel* (Klinger et al. 1982), and *Purposes of Vehicle Trips and Travel* (Roskin 1980). Based on these reports, the

number of driver trips during these hours of weekend nights in the five years was estimated to be 2,131 million. In the FARS data, the total number of single-vehicle driver fatalities in this period was 1,866. The estimated overall fatality rate is then 8,755 fatalities in 10 million trips. This rate compares with rates estimated by Summers and Harris (1978) for the broader class of all fatal accidents: 2.2 in 10 million for drivers with a BAL less than .1 and 73.9 in 10 million trips for drivers with BAL greater than .1. The accuracy of this estimate is not important to the sensitivity analysis since it serves to scale all conditional fatality rates but is factored out of any relative rates.

Given the two tables $\Pr(\mathbf{X})$ and $\Pr(\mathbf{X}|F)$, each having positive probabilities in all 1,920 classifications, cells of the six-way table $\Pr(F|\mathbf{X})$ were estimated by eqn (1). Unadjusted fatality rates over each margin, $\Pr(F|X_{i..})$, could be found by summing these cell probabilities weighted by corresponding probabilities from the table $\Pr(\mathbf{X})$. Except for inaccuracy in fitting the 1,920 nonzero cells, this would give results identical to using observed marginal frequencies of the FARS and NRBS data in the Bayes' equation. For example,

$$\Pr(F|\text{sex} = \text{male}) = \Pr(\text{sex} = \text{male}|F) \cdot \Pr(F)/\Pr(\text{sex} = \text{male}),$$

where $\Pr(\text{sex} = \text{male})$ is estimated by the proportion of males in the NRBS and $\Pr(\text{sex} = \text{male}|F)$ is estimated by that proportion in FARS. Table 8 shows these estimates obtained by summation over the six-way table (method A), and compares them with those taken directly from marginal frequencies of the two samples (method B). Discrepancies represent distortion introduced by the method of conditioning and averaging to create a table of 1,920 nonzero cells. Ratios of the smoothed fatality rates between categories of any variable in Table 8 are unadjusted and, in general, will not agree with those calculated from the odds ratios in Table 5 which have been adjusted for other covariables.

## 6. SENSITIVITY ANALYSIS: OBJECTIVES AND METHODS

Comparing FARS and NRBS tables of cell probabilities is an alternative to logistic regression for estimating the relative risk ratio for any two trip classifications $X_{ijk}$ and $X_{i',j',k'}$. The smoothing procedure to measure average effects over cell categories insures that all cells are nonzero and the risk ratio

$$r = \Pr(X_{ijk}|F) \cdot \Pr(X_{i',j',k'})/\Pr(X_{ijk})\Pr(X_{i',j',k'}|F) \qquad (2)$$

is always defined. Like the logistic regression model, this model can be used to predict how the overall fatality rate will change if the distribution of trip characteristics changes. The logistic regression model predicts how the fatality rate would change if the distribution of a single characteristic, such as driver BAL, were to change while the proportions of all other trip characteristics remain constant. In contrast, the sensitivity analysis preserves associations among the distributions of the six trip characteristics of the model, while changing one or two marginal distributions.

The procedure was to set a first marginal distribution at some extreme deviation from NRBS probabilities and then to change a second marginal distribution so as to recover the original observed fatality rate ($\Pr_0(F) = 8,755$ per million trips.) Changes to the first margin in step one of this procedure were made according to specific criteria for direction and distance as follows. In separate trials the probability in each category of the target margin was set close to zero or to one. Probabilities in other categories of the target margin were increased or decreased in proportion to their NRBS values at each of the 2K trials for a margin with K categories. In the case that a margin was dichotomous, only two trials were unique, since increasing the probability in the first category was then equivalent to decreasing probability in the second category. In each trial the probability in each target category was set close to zero or one, subject to constraints defined by the second step of the sensitivity analysis. The constraints were

M. A. CONNOLLY *et al.*

Table 8. Marginal fatality rates obtained from Bayes equation applied to individual cells of the smoothed six-way table (Method A) or to marginal frequencies of the original table (Method B). Rates are predicted number of fatalities in 10 million trips

| | Fatality rate by age | | | |
|---|---|---|---|---|
| Method | 20 years | 20–30 | 30–40 | 40+ |
| A | 10.51 | 10.36 | 7.88 | 5.35 |
| B | 10.47 | 10.31 | 7.80 | 5.43 |

| | Fatality rate by sex | |
|---|---|---|
| Method | Male | Female |
| A | 9.25 | 6.72 |
| B | 9.22 | 6.76 |

| | Fatality rate by number of occupants | | |
|---|---|---|---|
| Method | 1 | 2 | 3+ |
| A | 15.01 | 6.34 | 4.34 |
| B | 14.93 | 6.34 | 4.32 |

| | Fatality rate by BAL in mg/100 mL | | | | |
|---|---|---|---|---|---|
| Method | 0 | 0–.02 | .02–.05 | .05–.1 | .1–.15 |
| A | 1.53 | 1.17 | 1.28 | 11.81 | 84.82 |
| B | 1.17 | .11 | 2.31 | 8.39 | 43.02 |

| Method | .15–.2 | .2–.25 | .25–.3 | .3–.35 | .35+ |
|---|---|---|---|---|---|
| A | 388.8 | 1129. | 2093. | 2663.* | 2112.* |
| B | 252.9 | 607.0 | 859.0 | | |

| | Fatality rate by vehicle type | | | |
|---|---|---|---|---|
| Method | Small car | Mid-size | Large car | Cargo or recreational |
| A | 6.69 | 14.85 | 1.69 | 18.84 |
| B | | ( 7.48)† | | 20.94 |

| | Fatality rate by seat belt | |
|---|---|---|
| Method | Not used | Used |
| A | 12.28 | .85 |
| B | 11.74 | .65 |

*NRBS had no observations with BAL over .3 mg/100 mL.
†NRBS coding did not specify automobile size.

such that for each change to the first marginal distribution, a compensating change in a second marginal distribution would restore the average fatality rate to $Pr_0(F)$.

These disturbances to one or two marginal distributions were made so as to maintain the dependence structure among the distributions of the six trip characteristics, as observed in the NRBS data. Structure was defined in terms of the deviation of each cell probability from the independence model for a six-way table. In the first step, changes in the distribution of the first margin from that found in the NRBS, $(Pr(\mathbf{X}_{i\cdot\cdot}) \text{ to } Pr^*(\mathbf{X}_{i\cdot\cdot}))$, were imposed and the resulting changes in $Pr(F)$ were calculated as

$$Pr^*(F) = \sum_{ijk} (Pr(F|X_{ijk}) \cdot Pr^*(X_{ijk})). \qquad (3)$$

In the second step, $Pr(F)$ was held at the observed value and admissible changes in some second marginal distribution $Pr(X_{\cdot j\cdot})$ were determined that would yield that same probability. Because it was not possible to maintain the dependence structure or the NRBS

marginal distributions exactly, measures were devised to evaluate the extent of departure from stated goals.

The two steps of the sensitivity analysis shared one method for changing a six-way table subject to constraints on marginal distributions and dependence structure. This will be illustrated by a three-way table with marginal probabilities $\{\Pr(X_{i..})\}$, $\{\Pr(X_{.j.})\}$, and $\{\Pr(X_{..k})\}$ which has obvious generalization to higher-order tables. The notation $\{P_{ijk}\}$ will be used to indicate the set of cell probabilities $\{\Pr(X_{ijk})\}$. In similar notation, the cell probabilities for independence are $\pi_{ijk} = P_{i..}P_{.j.}P_{..k}$. The contribution of each cell to the $\chi^2$ test statistic for independence is $(\Delta_{ijk})^2$, where $\Delta_{ijk} = (P_{ijk} - \pi_{ijk})/(\pi_{ijk})^{1/2}$. When the distribution of the first margin is changed to $P_{i..}^*$ then $\{\pi_{ijk}^*\}$ is the new set of independence probabilities and $\pi_{ijk}^* = P_{i..}^* P_{.j.}P_{..k}$. The method for creating cell probabilities of a new marginal distribution is related to log linear models. Let the original table have an associated set of ratios $\lambda_{ijk} = P_{ijk}/\pi_{ijk}$. Then the new set of probabilities is

$$P_{ijk}^* = \lambda_{ijk}\pi_{ijk}^* / \sum_{ijk} (\lambda_{ijk}\pi_{ijk}^*). \tag{4}$$

In implementing this method, a set of margins, $\{P_{i..}^*\}$, $\{P_{.j.}\}$, and $\{P_{..k}\}$ is designated and the new set of probabilities $\{P_{ijk}^*\}$ are calculated according to (4). The true margins for this new table, $P_{i..}^{**} = \Sigma_{jk}P_{ijk}^*$, will be close to the designated set and the deviations from independence $(\Delta_{ijk})$ for each cell will also be similar between the old and new tables. This ratio method generalizes to changing one or more sets of margins from original NRBS values.

For changing a single margin, the ratio method is formally equivalent to a third approach based on changing the six-way table into a two-way array of the target margin by all other combinations, i.e. $\{P_{i..}\}$ by $\{P_{.jk}\}$. Then changes to the marginal distribution of the $i$th characteristics are distributed proportionally over the cells of $P_{.jk}$, i.e. $P_{ijk}^* = P_{.jk}P_{i..}/P_{i..}^*$. This method has some intuitive appeal, but if two margins ($i$ and $j$, for example) are to be changed, then the new set of two-way probabilities, $\{P_{ij.}^*\}$, must be specified. For any $\{P_{i..}^*\}\{P_{.j.}^*\}$ there are infinite choices for $\{P_{ij.}^*\}$.

Success of the ratio method in maintaining dependence structure was measured in several ways. Ideally, the sum of absolute relative differences

$$\sum_{ijk} |\Delta_{ijk} - \Delta_{ijk}^*|/\Delta_{ijk}$$

for the independence deviations should be small. However, because many of the $\Delta_{ijk}$ were close to zero this measure could be quite sensitive to the effects of a few cells. Alternate measures considered were the mean absolute difference relative to either the mean or the standard deviation of the original set $\{\Delta_{ijk}\}$. These measures of absolute relative difference were examined at each trial for qualitative comparison. However, the Pearson correlation coefficient between $\Delta$ and $\Delta^*$ appeared to be the best measure of average success in maintaining dependence structure at each trial.

The departure of any margin from its intended or designated values was measured by the Mahalanobis distance function, which is a weighted distance between two points in a multidimensional space. Let the vector $\mathbf{P}$ be the intended set of probabilities for the $i$th margin ($\mathbf{P}^*$ in the case that this margin was the target margin) and $\mathbf{P}^{**}$ the corresponding marginal probabilities of the normalized table after changing one or more margins. Then the Mahalanobis distance between $\mathbf{P}$ and $\mathbf{P}^{**}$ is

$$M_i = (\mathbf{P}^{**} - \mathbf{P})'\mathbf{V}^-(\mathbf{P}^{**} - \mathbf{P}), \tag{5}$$

where $\mathbf{V}^-$ is a generalized inverse of the multinomial covariance matrix for $\mathbf{P}$. The value of the quadratic form is invariant to the choice of generalized inverse. For comparison among all distances for margins of different dimensions, this distance was taken as a ratio of the largest distance from $\mathbf{P}$ to some element in the simplex $\{(1, 0, 0, \ldots), (0, 1, 0, \ldots), \text{etc.}\}$.

## 7. SENSITIVITY ANALYSIS: MEASURED OUTCOMES

In the first step of the sensitivity analysis, effects of changes to distributions of driver characteristics on the fatality rate were explored. The disturbances to the NRBS probabilities involved taking each marginal distribution toward its extreme values subject to the constraint that a change in any other marginal distribution was possible that would restore the original average fatality rate. For these changes in marginal distributions, the ratio method was used to calculate a new table of driver probabilities. These probabilities $\{P_{ijk}^*\}$ were entered in eqn (3) to calculate a new average fatality rate $\Pr^*(F)$ resulting from a change to a single characteristic of the driving population. This rate is reported, along with measures that would indicate the extent to which other marginal distributions and the NRBS dependence structure were maintained. Provided that the slippage of all margins from their intended value is small and that changes in dependence structure are small, results of changing a single margin while maintaining the dependence structure can be evaluated.

In the second step of the sensitivity analysis, changes were made to each other margin that would restore the overall average fatality rate ot $\Pr_0(F) = 8.755$ in 10 million trips after one marginal distribution had been set to a limiting distribution as described in the previous paragraph. For each limiting distribution in step 1, each of the 5 other marginal distributions may be changed to solve an equation that sets the average fatality rate to $\Pr_0(F)$. The point in the solution set which is reported is that distribution which

Table 9a. NRBS percent age distribution by state

| State | Age | | | |
|---|---|---|---|---|
| | Under 20 | 20–30 | 30–40 | 40+ |
| New Jersey | 16.9 | 45.0 | 12.1 | 25.0 |
| Oregon | 20.9 | 41.8 | 19.4 | 17.9 |
| Virginia | 19.8 | 35.6 | 22.8 | 21.8 |
| Wisconsin | 27.9 | 34.9 | 14.0 | 23.3 |
| California | 20.3 | 34.5 | 21.6 | 23.6 |
| Washington | 21.8 | 34.1 | 19.2 | 25.0 |

Table 9b. NRBS percent sex distribution by state

| State | Sex | |
|---|---|---|
| | Male | Female |
| California | 85.8 | 14.2 |
| Wisconsin | 85.3 | 14.7 |
| Virginia | 84.2 | 15.8 |
| Oregon | 80.6 | 19.4 |
| New Jersey | 79.8 | 20.2 |
| Washington | 77.0 | 23.0 |

Table 9c. NRBS percent of number of occupants distribution by state

| State | Number of occupants | | |
|---|---|---|---|
| | 1 | 2 | 3+ |
| Virginia | 33.7 | 21.8 | 44.6 |
| California | 35.8 | 34.5 | 29.7 |
| Washington | 34.1 | 36.4 | 29.6 |
| Oregon | 38.8 | 33.6 | 27.6 |
| New Jersey | 25.8 | 50.8 | 23.4 |
| Wisconsin | 41.1 | 36.4 | 22.5 |

Table 9d. NRBS percent BAL distribution by state (in mg/dL)

| State | BAL | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 0–0.02 | 0.02–0.05 | 0.05–0.10 | 0.10–0.15 | 0.15–0.20 | 0.20–0.25 | 0.25–0.30 |
| New Jersey | 23.5 | 56.3 | 5.9 | 9.2 | 2.5 | 2.5 | — | — |
| Virginia | 30.2 | 44.8 | 6.2 | 12.5 | 4.2 | 2.1 | — | — |
| Wisconsin | 51.7 | 15.0 | 15.0 | 12.5 | 5.8 | — | — | — |
| California | 60.3 | 18.4 | 7.8 | 8.5 | 4.3 | 0.7 | — | — |
| Washington | 62.7 | 13.4 | 12.7 | 7.5 | 3.1 | 0.3 | 0.3 | — |
| Oregon | 67.4 | 15.5 | 2.3 | 10.0 | 1.6 | 0.8 | 1.6 | 0.8 |

Table 9e. NRBS percent vehicle body type distribution
by state

| State | Vehicle type | |
|---|---|---|
| | Automobile | Cargo or recreational |
| New Jersey | 97.6 | 2.4 |
| Wisconsin | 91.5 | 8.5 |
| California | 91.2 | 8.8 |
| Washington | 87.0 | 13.0 |
| Virginia | 85.2 | 14.8 |
| Oregon | 82.1 | 17.9 |

Table 9f. NRBS percent using seat belt
by state

| State | No belt | Belt used |
|---|---|---|
| Oregon | 61.4 | 38.6 |
| California | 63.8 | 36.2 |
| New Jersey | 69.8 | 30.2 |
| Wisconsin | 69.9 | 30.1 |
| Washington | 71.6 | 28.4 |
| Virginia | 82.6 | 17.4 |

is closest to the NRBS distribution of the trip characteristic. The measure of closeness is the Mahalanobis distance defined by eqn (5). The rules for setting limit distributions for the first margin (step 1) and then choosing a solution for each other margin that restores $Pr_0(F)$ (step 2) are described in greater detail in Appendix A.

The key results of steps 1 and 2 are new distributions of each of the six trip characteristics. These may be compared with the NRBS distributions of Table 2 or the distributions within the six states listed in Table 9. The distributions of driver age and sex, number of occupants, driver BAL, vehicle body type, and seat belt use show how the marginal distributions vary within arbitrary classes of athe NRBS sample.

## 8. SENSITIVITY ANALYSIS RESULTS

Changes to single margins and resulting changes in $Pr(F)$ (step 1 results) are reported in Table 10. The columns show the lower and upper limits of changes to the probability in each target category. The key results of those changes are the new overall average fatality rates reported as percent change in $Pr(F)$. The correlation of the delta or chi-statistics over the cells of the old and new six-way tables indicate the degree of success in maintaining dependence structure. The standardized Mahalanobis distance values for the worst-fit second margin measure success in imposing a new set of marginal distributions. The worst-fit second margin indicates relatively large dependencies in the table of probabilities. The last column indicates the type of constraint (as described in Appendix A) active at the reported limit.

The direction of $Pr(F)$ changes reported in Table 10 corresponds to the relative trip

Table 10. Results of sensitivity analysis step 1: Changes to average fatality rate when single margins are set to limit distributions

| Target margin (NRBS probabilities) | Target category | Probability in target category | Correlation of $\Delta_{ijk}$ | Worst-fit margin | Mahalanbis distance and worst fit | % Change in Pr($F$) | Limit |
|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | |
| (.213) | <20 | .131 | .96 | seat belt | .00013 | -1.98 | * |
| | | .290 | .97 | seat belt | .00011 | 2.05 | * |
| (.369) | 20–29 | .221 | .95 | seat belt | .00014 | -4.18 | † |
| | | .490 | .96 | seat belt | .00009 | 3.60 | * |
| (.182) | 30–39 | .062 | .92 | number occ. | .00030 | 1.58 | †† |
| | | .317 | .94 | number occ. | .00037 | -1.55 | * |
| (.236) | 40+ | .134 | .96 | seat belt | .00010 | 5.28 | * |
| | | .337 | .96 | seat belt | .00010 | -5.06 | † |
| **Vehicle type** | | | | | | | |
| (.178) | small car | .004 | .85 | sex | .00002 | 5.12 | |
| | | .294 | .95 | sex | .00001 | -3.32 | † |
| (.342) | medium car | .299 | .98 | sex | .00000 | -4.42 | † |
| | | .720 | .82 | sex | .00017 | 40.04 | § |
| (.393) | large car | .196 | .91 | age | .00003 | 26.22 | † |
| | | .432 | .98 | age | .00000 | -5.21 | † |
| (.087) | truck or van | .057 | .97 | sex | .00001 | -4.56 | † |
| | | .220 | .91 | sex | .00029 | 20.39 | § |

| Variable | | value | | compensated by | Pr(F) | % change | note |
|---|---|---|---|---|---|---|---|
| **Number of occupants** | | | | | | | |
| one occupant | (.347) | .305 | .98 | sex | .00001 | -4.45 | † |
| | | .645 | .87 | sex | .00048 | 32.68 | § |
| 2 occupants | (.358) | .007 | .72 | age | .00069 | 15.22 | † |
| | | .465 | .96 | age | .00006 | -4.51 | |
| 3 or more | (.295) | .006 | .74 | age | .00057 | 20.75 | † |
| | | .354 | .97 | age | .00002 | -4.12 | † |
| **Sex of driver** | | | | | | | |
| male | (.808) | .700 | .94 | number occ. | .00016 | -3.01 | †† |
| | | .908 | .91 | number occ. | .00014 | 2.98 | †† |
| **Seatbelt used** | | | | | | | |
| yes | (.308) | .135 | .90 | age | .00040 | 22.81 | ‖ |
| | | .348 | .99 | age | .00001 | -5.10 | † |
| **Blood alcohol level** | | | | | | | |
| BAL=0 | | .451 | .97 | number occ. | .00015 | 26.28 | †† |
| | | .557 | .99 | number occ. | .00002 | -5.38 | † |
| [0 - .02)mg/dL | | .128 | .96 | number occ. | .00016 | 9.36 | †† |
| | | .285 | .98 | number occ. | .00003 | -5.09 | † |
| [.02 - .05)mg/dL | | .000001 | .84 | number occ. | .00020 | 12.87 | |
| | | .18 | .98 | number occ. | .00003 | -4.80 | † |
| [.05 - .1)mg/dL | | .000001 | .87 | number occ. | .00006 | -2.81 | |
| | | .212 | .92 | number occ. | .00018 | 5.21 | †† |
| [.1 - .15)mg/dL | | .014 | .99 | seat belt | .00000 | -5.23 | † |
| | | .055 | .96 | seat belt | .00005 | 30.79 | † |

*The seatbelt use distribution to compensate does not bring Pr (F) to within 1% of original value for more extreme change to first margin.
†A more extreme change cannot be compensated by any sex of driver distribution.
††As (*) for number of occupants.
§As (*) for sex of driver.
‖As (*) for age of driver.

fatality rates for different portions of the driving population. Thus, for example, in Table 8 it may be seen that male drivers have a higher average fatality rate than females. In Table 10, an increase in the proportion of male drivers from .808 to .908 increases the average fatality rate 2.98%. The degree of change in $\Pr(F)$ as $\Pr(X_{i\cdot\cdot})$ is changed depends in part on the differences among the conditional fatality rates $(\Pr(F|X_{i\cdot\cdot}))$ of Table 8. For example, Table 8 reports a larger difference between conditional fatality the two seat belt use categories than for the two sex categories. The range in $\Pr(F)$ possible for changing the seat belt use probabilities is correspondingly larger in Table 10 than the range for changing the sex distribution. The range in $\Pr(F)$ for all changes to the age distribution is from $-5.06\%$ to $5.28\%$. This is a small range compared to results for other margins, and it is consistent with relative similarity among conditional fatality rates reported for the age categories in Table 8.

Constraints of preserving the dependence structure and solving for step 2 solutions to restore $\Pr_0(F)$ may limit the percent change in $\Pr(F)$ reported in Table 10. Many changes to a single marginal distribution are possible that greatly increase the average fatality rate. Changes to the vehicle distributions increase $\Pr(F)$ by 20.39%, 26.22%, or 40.0%. Changes in the number of occupants distribution can increase the fatality rate by 15.20% or 20.75%. Large increases are also possible as a result of changes to the BAL distribution. The constraint that step 2 solutions must exist allows no change that reduces the average fatality rate by much more than 5%. Many changes that reduce $\Pr(F)$ are noted to have such a constraint imposed by the sex of driver distribution.

Solutions for all margins to restore $\Pr_0(F)$ may not exist if step 1 changes greatly increase $\Pr(F)$. If the proportions of drivers with extremely high BAL (.3 and over) are increased, then $\Pr(F)$ may be increased by as much as 40%, calling for a proportion near 1 of drivers in the age 40 and over category to compensate for this increase. Some changes, such as an increase in the proportion of large cars or moderately high BAL categories, have a limited range because a step 2 solution calls for all drivers to be female. Sex tends to be a limiting second margin because marginal fatality rates are similar for male and female drivers.

In a few cases constraints do not come into play. The probabilities in some categories may be reduced essentially to zero within the constraints of the sensitivity analysis. Such classes are vehicles with two or with three or more occupants. Changing these categories to zero increases the $\Pr(F)$ by 15.22% and 20.75%, respectively. Other categories that may be reduced to zero probability are the BAL levels .02–.05 mg/dL and .05–.1 mg/dL. Also, the probabilities for BAL categories .3 to .35 and .35 and over may be reduced to zero, although this represents only a small change from original smoothed NRBS probabilities.

Some constraints in Table 10 are not based on the existence of step 2 solutions but on dependence structure. If a single marginal distribution is set to be very different from the NRBS values, causing a large change in $\Pr(F)$ and requiring a large adjustment in any other margin to compensate, then constraints of the fitting procedure may not be met in step 2. An attempt to set two margins far from NRBS values and maintain the NRBS dependence structure may cause all margins in the new $\Pr^*(X)$ table to be different from those entered. The new $\Pr(F)$ will not be within 1% of $\Pr_0(F)$. For example, there is a dependence structure constraint for a proportion of teen-aged drivers below .131 or over .290 when the compensating second margin is seat belt use. Past this range, the fitting procedure [eqn (4)] will cause margins to shift from set values and the new $\Pr(F)$ [evaluated by eqn (3)] will not be within 1% of $\Pr_0(F)$.

In Table 10 are four cases where the joint distribution of driver age and seat belt use limit permissible changes to the age distribution. These cases are the range for increasing or decreasing the proportion of teen-aged drivers, for increasing the proportion of drivers in their 20s or 30s, and for decreasing the proportion of drivers 40 and over. In another dependence structure limit of Table 10, the degree to which vehicle type and sex of driver may be simultaneously changed limits increase in proportions of medium-sized cars or of trucks and vans in step 1. Also, the proportion of cars with one occupant

cannot be increased above .645 if the sex of driver distribution is to be the compensating second margin.

The column reporting Mahalanobis distance [eqn (5)] indicates that nontarget margins shifted very little from NRBS values when a single margin was modified. Some dependencies in the six-way table indicated by the column labeled worst-fit margin correspond to conditioning steps of the smoothing procedure. For example, seat belt use was conditioned on driver age, and BAL was fit within age by sex by number of occupant categories. The correlation between the cell $\Delta_{ijk}$ values or chi-statistics of the original and modified tables tended to decrease as the degree of modification to the target margin increased. Other limits on changes to the target margin prevented this correlation from falling below .72 and for most changes the correlation exceeded .95. Generally, the dependence structure, as measured by the $\chi^2$ test for six-way independence, was maintained.

Extremity of changes to marginal distributions in Table 10 may also be compared to Table 9, which reports these distributions within the six states of the NRBS sample. Thus, Table 10 indicates that the proportion of teen-aged drivers was allowed to range between 0.131 and 0.290. The actual state-to-state variation in this proportion was 0.169–0.279. The two ranges are not entirely comparable since imposed changes were proportional to NRBS values and probabilities within states need not show such patterns. Changes to two marginal distributions that yield original average fatality rates are reported in Tables 11a–e. In these examples, changes to target category caused large changes in Pr($F$). Compensating distributions reported in Tables 11a–11e are those that limit changes to target category. The proportional distribution over the four age categories is listed in Table 11a. Among age categories (Table 9a) the observed range over the states for 30–39 year olds, .121 to .228, is smaller than the step 1 limits of .062 to .317. Even these wide changes to the age distribution resulted in a small range of Pr*($F$) [from +1.58% to −.1.55% of Pr$_0$($F$)] compared to changes in other categories of the age distribution.

In Table 9e, the proportion of trucks and vans among all vehicles ranged from .024 to 0.179. This compares with a range in Table 10 of .057–.220, which corresponds to Pr*($F$) that were −4.56% or +20.39% of Pr$_0$($F$). Observed proportions of male drivers varied among states from .770 to .858 (Table 9b). Step 1 limits of .700 to .908 caused Pr*($F$) to be 3.01% lower or 2.98% higher than Pr$_0$($F$). Seat belt use rate across states ranged from .174 to .386. In the sensitivity analysis belt use rates of .134 or .348 were calculated to increase Pr($F$) by 22.81% or decrease Pr($F$) by 5.10%. Table 9d reports observed proportions of drivers with respect to the lower eight BAL categories (0 to .3 mg/dL). The proportion of drivers with BAL = 0 ranged from .235 to .674 in the different states. Table 12 examines changes in second margins to compensate shifts in BAL distribution. The range reported in Table 12a is .451 to .557, corresponding to Pr*($F$) that is +26.28% or −5.58% of Pr$_0$($F$). Proportions in the BAL category .15 − .2 mg/dL ranged from 0 to .26. In Table 12a the range is .004 to .012, predicting a Pr($F$) that is −5.22% or +31.10% of Pr$_0$($F$).

Only a selection of step 2 results are presented in Tables 11a–e and in Tables 12a–d. Step two results omitted from these tables are in Appendix B. Tables 11a–e report results compensating for one target category in each non-BAL margin in Table 10. Tables 12a–d report step 2 results for changes in driver BAL distribution compensated by changes to each of the five other margins. Step 1 target categories chosen for Tables 11a–c are those that could be set to cause the widest range of Pr($F$) values among all changes to the respective target distribution. No selection for category was needed for the dichotomous sex and seat belt use margins (Tables 11d and e). Step 2 solutions reported in Tables 11a–e are limiting second margins that determined at least one endpoint of the range for corresponding first-margin categories. In Table 10, step 1 changes are indicated by probabilities in the target category. Examples in Table 11 show corresponding probabilities in all categories of the target margin.

The age distributions represented in Table 11a are limits for changes in the pro-

Tables 11a–11e. Changes to two marginal distributions which yield original average fatality rate.
In these examples changes to target category caused large changes in Pr(F).
Compensating distributions reported here are those that limit changes to target category

Table 11a. Changes to driver age distribution. Target category is ages 40 and over. Sex of driver (*) and seatbelt use (†) are limiting

| Categories of first margin | Probabilities | | | Categories of second margin | Probabilities | | |
|---|---|---|---|---|---|---|---|
| | NRBS | Decrease target | Increase target | | NRBS | Compensate decrease | Compensate increase |
| <20 | .213 | .241 | .185 | Male driver | .808 | .637 | .994 |
| 20–29 | .369 | .418 | .320 | Female | .192 | .363 | .006 |
| 30–39 | .182 | .207 | .158 | Seat belt used | .308 | .340 | .274 |
| 40+ | .236 | .134 | .337 | Seat belt not used | .672 | .660 | .726 |

Table 11b. Changes to vehicle type distribution. Target category is medium-size car. Sex of driver (*) and number of occupants (†) are limiting

| Categories of first margin | Probabilities | | | Categories of second margin | Probabilities | | |
|---|---|---|---|---|---|---|---|
| | NRBS | Decrease target | Increase target | | NRBS | Compensate decrease | Compensate increase |
| Small car | .178 | .189 | .076 | Male driver | .808 | .975 | .003 |
| Medium car | .342 | .300 | (.720) | Female driver | .192 | .025 | .977 |
| Large car | .393 | .418 | .167 | 1 occupant | .347 | .384 | .106 |
| Truck or van | .087 | .093 | .037 | 2 occupants | .358 | .360 | .360 |
| | | | | 3 or more | .295 | .256 | .534 |

Table 11c. Changes to number of occupants distribution. Target category is one occupant. Sex of driver (*) is limiting

| Categories of first margin | Probabilities | | | Categories of second margin | Probabilities | | |
|---|---|---|---|---|---|---|---|
| | NRBS | Decrease target | Increase target | | NRBS | Compensate decrease | Compensate increase |
| 1 occupant | .347 | .305 | .645 | Male driver | .808 | .978 | .110 |
| 2 occupants | .358 | .381 | .195 | Female driver | .192 | .022 | .890 |
| 3 or more | .295 | .314 | .160 | | | | |

Table 11d. Changes to sex of driver distribution compensated by changes to number of occupants (†) distribution

| Categories of first margin | Probabilities | | | Categories of second margin | Probabilities | | |
|---|---|---|---|---|---|---|---|
| | NRBS | Decrease target | Increase target | | NRBS | Compensate decrease | Compensate increase |
| Male driver | .808 | .700 | .908 | 1 occupant | .347 | .382 | .316 |
| Female driver | .192 | .300 | .092 | 2 occupants | .358 | .360 | .360 |

*Indicates limit based on eqn (5).
†Indicates limit due to dependence structure.

Table 11e. Changes to seat belt use distribution compensated by changes to sex of driver (*) and age of driver (†) distributions

| Categories of first margin | Probabilities | | | Categories of second margin | Probabilities | | |
|---|---|---|---|---|---|---|---|
| | NRBS | Decrease target | Increase target | | NRBS | Compensate decrease | Compensate increase |
| Seat belt used | .308 | .134 | .348 | Male driver | .808 | .157 | .994 |
| Seat belt not used | .692 | .866 | .652 | Female | .192 | .843 | .006 |
| | | | | <20 yrs. | .213 | .206 | .216 |
| | | | | 20–29 | .369 | .060 | .460 |
| | | | | 30–39 | .182 | .180 | .180 |
| | | | | 40+ | .236 | .554 | .144 |

*Indicates limit based on eqn (5).
†Indicates limit due to dependence structure.

portion of drivers 40 and over. The probabilities indicated for these limits (.134 and .337) correspond to entries in Table 10. The age distributions in Table 11a are determined by the probability in the target category and the criterion that probabilities in other categories are adjusted in proportion to corresponding NRBS values. Table 11a shows proportions of male drivers (.637 and .994) that would correspond to the original average $\Pr(F)$ if the proportion of drivers age 40 and over were decreased or increased, respectively. Seat belt use rates are also shown that would compensate for these changes to the oldest age category. Notations in Table 11a indicate that sex of driver was a limiting second margin based on eqn (5) and that seat belt was limiting due to dependence between marginal distributions.

In Tables 11a and e, sex of driver is a limiting second distribution. Increasing the proportion of drivers who are male compensates for changes that decrease $\Pr(F)$. Changes to single margins, which, with an entirely male driving population would maintain $\Pr_0(F)$, are an increase in the proportion of drivers over age 40 from .236 to .337, an increase in the proportion of medium-sized cars from .342 to .720, a decrease in the proportion of single-occupant trips from .347 to .305, or an increase in seatbelt use from .308 to .348.

Tables 12a–d report sample changes to the distribution of driver BAL along with step 2 solutions for each of the five other margins. The reported BAL categories are the six lowest levels, since effects in the tail of the BAL distribution are as likely to be artifacts of the smoothing procedures rather than representative of the driving population. As in Table 11, the reported point solution for distribution of age, vehicle type, and number of occupants which restore $\Pr_0(F)$ were found by a grid search over the solution set to find a point near the NRBS distribution. Due to the grid size (.02) and use of the Mahalanobis formula to evaluate distance from NRBS values, probabilities in particular categories (30- 40-year-olds, small cars, and two occupants) are identical in many solution sets of Tables 12a, b, and c. By comparison with Table 8, it may be seen that those columns that show the smallest shift from NRBS values have marginal fatality rates closest to the overall $\Pr_0(F)$.

The types of limits in effect for these changes to six BAL categories are reported in Table 10. The widest changes to $\Pr(F)$ are possible when probabilities in the fifth and sixth BAL levels are shifted to limit distributions. The range of $\Pr(F)$ is also quite wide for a shift of $\Pr(\text{BAL} = 0)$. In Table 12 it may be seen that compensating distributions which are most different from NRBS margins are solutions for step 1 changes to probabilities in the fifth, sixth, and first BAL categories.

## 9. IMPLICATIONS FOR DRIVING RISKS

This report evaluates marginal and adjusted effects of risk factors for single-driver fatalities on weekend nights. Results of logistic regressions predict effects of changing

Table 12a. Changes in driver age distribution to compensate BAL distribution shifts

| BAL catgory from NRBS to new | | | (NRBS) | Age distribution | | | |
|---|---|---|---|---|---|---|---|
| | | | | <20 .213 | 20–30 .369 | 30–40 .182 | 40+ .236 |
| 0 BAL | .526 | .451 | | .160 | .220 | .182 | .438 |
| | | .557 | | .220 | .462 | .180 | .138 |
| 0–.2 mg/dL | .240 | .128 | | .200 | .194 | .180 | .426 |
| | | .285 | | .240 | .440 | .182 | .138 |
| .02–.05 mg/dL | .130 | .000 | | .208 | .180 | .180 | .432 |
| | | .180 | | .220 | .448 | .180 | .152 |
| .05–.1 mg/dL | .077 | .000 | | .218 | .420 | .180 | .182 |
| | | .212 | | .149 | .360 | .180 | .311 |
| .1–.15 mg/dL | .020 | .014 | | .220 | .460 | .183 | .137 |
| | | .055 | | .009 | .220 | .180 | .591 |
| .15–.2 mg/dL | .005 | .004 | | .220 | .460 | .183 | .137 |
| | | .012 | | .180 | .018 | .180 | .622 |

Table 12b. Changes in vehicle type distribution to compensate BAL distribution shifts

| BAL category from NRBS to new | | | (NRBS) | Vehicle type | | | |
|---|---|---|---|---|---|---|---|
| | | | | Small car .178 | Medium car .342 | Large car .393 | Truck or van .087 |
| 1 BAL | .526 | .451 | | .180 | .340 | .445 | .035 |
| | | .557 | | .180 | .360 | .360 | .100 |
| 0–.2 mg/dL | .240 | .128 | | .180 | .340 | .444 | .036 |
| | | .285 | | .180 | .360 | .361 | .099 |
| .02–.05 mg/dL | .130 | .000 | | .180 | .340 | .444 | .036 |
| | | .180 | | .180 | .360 | .362 | .098 |
| .05–.1 mg/dL | .077 | .000 | | .178 | .340 | .380 | .102 |
| | | .212 | | .180 | .340 | .414 | .066 |
| .1–.15 mg/dL | .020 | .014 | | .180 | .360 | .360 | .100 |
| | | .055 | | .180 | .300 | .512 | .008 |
| .15–.2 mg/dL | .005 | .004 | | .180 | .360 | .360 | .100 |
| | | .012 | | .180 | .300 | .513 | .007 |

Table 12c. Changes in number of occupants distribution to compensate BAL shifts

| BAL category from NRBS to new | | | (NRBS) | Number of occupants | | |
|---|---|---|---|---|---|---|
| | | | | 1 .347 | 2 .358 | 3 or more .295 |
| 0 BAL | .526 | .451 | | .246 | .360 | .394 |
| | | .557 | | .398 | .360 | .242 |
| 0–.2 mg/dL | .240 | .128 | | .258 | .360 | .382 |
| | | .285 | | .387 | .360 | .253 |
| .02–.05 mg/dL | .130 | .000 | | .258 | .360 | .382 |
| | | .180 | | .384 | .360 | .256 |
| .05–.1 mg/dL | .077 | .000 | | .375 | .360 | .265 |
| | | .212 | | .295 | .360 | .344 |
| .1–.15 mg/dL | .020 | .014 | | .392 | .360 | .248 |
| | | .055 | | .141 | .360 | .499 |
| .15–.2 mg/dL | .005 | .004 | | .392 | .360 | .248 |
| | | .012 | | .151 | .360 | .489 |

Table 12d. Changes in driver sex and seat belt use distributions to compensate BAL shifts

| BAL Category from NRBS to new | | | (NRBS) | Driver's sex | | Seat belts | |
|---|---|---|---|---|---|---|---|
| | | | | Male | Female | Used | Not used |
| | | | | .088 | .192 | .308 | .692 |
| 0 BAL | .526 | .451 | | .388 | .612 | .397 | .603 |
| | | .557 | | .999 | .001 | .264 | .736 |
| 0–.2 mg/dL | .240 | .128 | | .432 | .568 | .393 | .607 |
| | | .285 | | .9999 | .0001 | .268 | .732 |
| .02–.05 mg/dL | .130 | .000 | | .461 | .539 | .393 | .607 |
| | | .180 | | .9997 | .0003 | .270 | .730 |
| .05–.1 mg/dL | .077 | .000 | | .895 | .105 | .288 | .712 |
| | | .212 | | .618 | .382 | .343 | .657 |
| .1–15 mg/dL | .020 | .014 | | .9998 | .002 | .266 | .734 |
| | | .055 | | .003 | .997 | .484 | .516 |
| .15–.2 mg/dL | .005 | .004 | | .9999 | .0001 | .266 | .734 |
| | | .012 | | .0007 | .9993 | .488 | .512 |

a single factor with other effects held constant, while the sensitivity analysis predicts the fatality rate if a single factor and any associated characteristics are changed. In this section the implications of statistics reported in several tables are compared. The median age in FARS is 23, three years less than the NRBS median (Table 2.). This indicates a risk for younger drivers that is also shown by the results of the logistic regression (Table 4), which predict a relative risk of .80 for 25-year-olds compared to 20-year-olds (or for any other five-year interval). The categorical logistic regression (Table 5) yields risks relative to teen-aged drivers of .59, .28, and .18 for older drivers in age groups 20–30, 30–40, and 40+, respectively. The corresponding relative risk estimates taken directly from marginal rates (Table 8), are .98, .74, and .52. This suggests that older drivers may have more risk due to other factors, such as high BAL, but that youth alone is also a risk. The sensitivity analysis predicts that increasing or reducing the proportion of teen-aged drivers reduces or increases Pr(F) by approximately 2% (Table 10). Changes to the lowest risk age category of 40 or older can change Pr(F) by ±5%.

Considering the risks for male drivers, the proportions in the FARS and NRBS samples (85.4% and 81.1%, respectively) indicate slightly more risk for male drivers. The logistic regressions show that the relative risk for women compared to men is 1.34, suggesting weakly that male drivers have more high-risk behaviors or factors, but in the absence of such factors females have more risk. The marginal odds ratio for women compared to men is .73 (Table 8). An increase or decrease in the proportion of males causes Pr(F) to increase or decrease by 3% (Table 10).

The proportion of cars with one occupant in FARS is 59.5%. In NRBS it is 34.9%. The logistic regression suggests that the relative risk for a fatality for two occupants compared to one is .50 and for three or more is .27. Corresponding relative risks estimated from marginal frequencies are .42 and .29 (Table 8). Thus, it appears that in this population drivers alone are at greater risk for a driver fatality than are those with passengers. The sensitivity analysis shows an increase in Pr(F) of 32.68% if the proportion of cars with one occupant is increased from .347 to .645 (Table 10).

The sources for vehicle body type information were less reliable than for other trip characteristics. The proportion of trips in light trucks or vans is higher in the FARS sample (24.1%) than in NRBS (11.2%). The logistic regression estimate of relative risk for nonauto compared to automobile is 1.99. Marginal rates correspond to a relative risk of 2.52 (Table 8). In the sensitivity analysis, an increase in the proportion of trucks and light vans from .087 to .220 increases Pr(F) by 20.39% (Table 10). There is evidence for higher risk to light trucks and vans. In consideration of changing demand for non-traditional vehicles and the detail available in the FARS data concerning automobile size, it would be interesting to have results of a more current survey showing similar coding to the FARS data for vehicle type.

Seat belt use is evidently protective, as the proportion using seat belts in the NRBS sample is 27.1% compared to 2% in the FARS sample. The FARS sample is only 56% complete for this variable. A logistic regression that includes missing belt use as a third category leads to relative risk estimates of .24 and 6.05 for seat belt used or not reported compared to seat belt not used. The marginal relative risk estimate based on nonmissing values is .23 (Table 8). Thus, it appears that seat belt use is highly protective and is also associated with other risk-reducing characteristics. The sensitivity analysis predicts that an increase in seat belt use from .308 to .348 would reduce $Pr(F)$ by 5.1% (Table 10).

There is a large difference in the BAL of the two samples. On weekend nights the median BAL of drivers is 0, the median of the FARS sample is .17 gm/dL. Logistic regression estimates of relative risk for BAL were .09, 1.60, 6.82, and 31.2 for $0 < \text{BAL} < .02$ gm/dL, $.02 \le \text{BAL} < .05$, $.05 \le \text{BAL} < .10$, and $.10 \le \text{BAL} < .15$. Corresponding relative risks based on marginal frequencies were .09, 1.97, 7.17, and 36.8, respectively (Table 8). The apparent protective effort of low BAL levels might be attributable to nonresponse bias or coding round-off. In the sensitivity analysis, proportional adjustment to the nontarget categories may have influenced predicted changes in $Pr(F)$. Increasing the proportion of drivers in any BAL levels under .05 decreased $Pr(F)$. Slight increases in any of the BAL levels above .1 greatly increased $Pr(F)$.

## 10. DISCUSSION

This paper presents methods to combine data from various sources to evaluate relative risk when no appropriate case-control or prospective studies exist. Its application requires that distribution data for relevant risk variables has been reported for the case population (traffic fatalities) and the population at risk (drivers). Compromises must be made in choosing and coding variables for the risk model based on availability and comparability of the data. Our study was based primarily on two samples from the FARS and NRBS records, augmented with some data from DOT surveys.

Initial indications of relative risk to different classes based on comparison of means and frequencies within the FARS and NRBS samples were not contradicted by subsequent analysis. Logistic regression methods were used to rank the importance of the risk variables age, sex, number of occupants, BAL, vehicle type, and seat belt use. Regression estimates and their standard errors may be used to make inferences about which risk factors are most significant for modifying the risk of a fatality within the context of other risk factors. The important assumption of the logistic regression model, that fatalities are from the population represented by the controls, is violated in our application, so that use of this standard statistical procedure for our nonstandard sample might suggest more certainty than is reasonable.

If all data on trip characteristics and fatalities were from a single retrospective study, then the paper by Breslow and Powers (1978) justifies use of the prospective logistic regression model to estimate odds ratios. If a survey existed that was carefully designed to complement the FARS data then the combined data might be considered a synthetic retrospective design and the prospective logistic regression model would be a useful tool for analysis. However, because of the poor match between populations sampled by the FARS and NRBS data, the coefficients estimated in this report are used only to indicate the direction and magnitude of relative risks subject to comparability between the samples.

The sensitivity analysis avoids the stochastic assumptions of logistic regression. As was demonstrated, more than two sources for data may be incorporated in the model. DOT sources were used to model automobile weight classes and to estimate an overall trip fatality rate. The principles are to explore the range of models that may be consistent with observed data and to exclude distributions of risk factors that have zero probability rather than to emphasize confidence intervals about relative risk estimates.

All sources for information about fatalities, the population at risk, and average fatality rates are used to estimate the table of fatality rates conditioned on specific sets of trip characteristics. This is the data summarization step of the procedure. The table

could be used to predict fatality rates for new populations or for samples selected from the NRBS data, provided that these populations can be summarized in the same six-way table of frequencies. We chose to apply these fatality rates to hypothetical populations representing simple changes to the NRBS sample.

Combining information from various sources in this way is a less-effective means to evaluate risk factors than would be a designed retrospective or prospective study. However, it makes use of data that is readily available and is much less expensive and time-consuming for such rare events as traffic fatalities that occur to a transient and ill-defined risk group. A primary weakness in this study is the low correspondence between the populations and variables reported in the FARS and NRBS data. The NRB Survey provided quality data for relevant descriptors of a population that was carefully defined with respect to time of driving and geographic sampling scheme. However, the survey was intended to be preliminary. It is a very small fraction of the target population and is also more than 10 years old. A current survey of similar type would be very useful and, combined with an analysis such as this, would be more feasible than a case-control study of similar scale.

## REFERENCES

Breslow, N., Powers, W. Are there two logistic regressions for retrospective studies? *Biometrics.* **34**:100–105; 1978.
Borkenstein, R. F. *et al.* The role of the drinking driver in traffic accidents (*The Grand Rapids Study*). Bloomington, IN: Department of Police Administration, Indiana University; 1964.
Borkenstein, R. F. Historical perspective: North American traditional and experimental response. *J. Stud. Alcohol, Proc. N. Am. Conf. Alcohol Highway Safety, J. Stud. Alcohol.* 10(suppl.):3–12; 1985.
Highway Statistic Division, Office of Highway Planning, Federal Highway Administration. Highway Statistics 1976 NTIS Rep No. PB 222479/AS; Highway Statistics 1977 NTIS Rep No. PB 296097; Highway Statistics 1978 NTIS Rep No. PB 80-126-808; Highway Statistics 1979 NTIS Rep No. PB 81-127-367; Highway Statistics 1980 NTIS Rep No. PB 82-142027; Highway Statistics 1981 NTIS Rep No. PB 83-128082.
Klinger, D., Kusmyak, J. R., Liss, F. *Household travel: Report* 9, 1977 *NPTS Rep No.* FHWA/PL/82/001. NTIS; 1982.
Kusmyak, J. R. *Vehicle occupancy: Report* 6, 1977 *NPTS.* NTIS Rep. No. FHWA/PL/81/012. NTIS; 1981a.
Kusmyak, J. R. *Household vehicle utilization: Report* 5, 1977 *NPTS.* NTIS Rep No. FHWA/PL/81/011. NTIS; 1981b.
National Highway Traffic Safety Administration, National Center for Statistics and Analysis. *Fatal accident reporting system.* NTIS Rep No. DOT HS-806-251. NTIS; 1981.
Roskin, M. E. *Purposes of vehicle trips and travel: Report* 3, 1977 *NPTS.* NTIS Rep No. FHWA/PL/81/001. NTIS; 1980.
Simpson, H. M. Human related risk factors in traffic crashes: Research needs and opportunities. *J. Stud. Alcohol.* 10(Suppl.):32–39; 1985.
Summers, L. G.; Harris, D. H. National Center for Statistics and Analysis. *The general deterrence of driving while intoxicated, volume* 1, *system analysis and computer-based simulation.* NTIS Rep No. DOT HS-803-582. NTIS; 1978.
Wolfe, A. C. 1973 *U.S. national roadside breathtesting survey: Procedures and results.* Prepared for the Office of Alcohol Countermeasures, DOT. Highway Safety Research Institute, Ann Arbor, Michigan; 1974.

## APPENDIX A: LIMITS FOR CHANGING A SINGLE MARGIN (STEP 1)

Limits to changes in single margins were determined as follows. A change in any single margin defines a range of valid probabilities for each other margin that will maintain $Pr(F)$ at the original value. If any margin other than the first is changed, then the equation that defines the first margin, $Pr(X_{i \cdot \cdot})$, is

$$Pr_0(F) = \sum_i Pr(F|X_{i \cdot \cdot}) \cdot Pr(X_{i \cdot \cdot}), \qquad (A1)$$

where $Pr_0(F) = 8.755$ in 10 million trips (DOT), and where

$$Pr(F|X_{i \cdot \cdot}) = \sum_{jk} (Pr(F|X_{ijk}) \cdot Pr^*(X_{ijk})). \qquad (A2)$$

For example, if the age distribution of drivers has been changed, the equation to define seat belt use probabilities is

$$Pr(F|\text{belt use}) \cdot Pr(\text{belt use}) + Pr(F|\text{no belt}) \cdot Pr(\text{no belt}) = 8.755 \times 10^7,$$

where $Pr(F|\text{belt use})$ is calculated from eqn (A2) using the new set of cell probabilities $Pr^*(X_{ijk})$ that resulted from changing the age probabilities from $P_{i \cdot \cdot}$ to $P^*_{i \cdot \cdot}$. The conditional fatality rates, $\{Pr(F|X_{ijk})\}$ estimated

from smoothed data are constants of the sensitivity analysis. Equation (A1) is then solved to find the Pr(belt use) that restores the original fatality rate after the age margin is changed. Note that the original (NRBS) age distribution gives an equation that is solved by the original belt use distribution.

As each target category is changed, eqn (A2) is examined for each other margin. Changes to the first margin beyond the limits reported in Tables 10 through 12 lack proper eqn (A2) solutions for at least one other margin. Reported step one results for each target category of each margin are the limit probability in the target category, the new $Pr^*(F)$, and, in cases that the limit probability is not 0 or 1, the second margin that first fails to have a solution to eqn (A1).

In step 2 of the sensitivity analysis, complementary changes are made in two marginal distributions. A first margin is set to limits as described in the preceding paragraph and then a second margin is changed so that the original average fatality rate, $Pr_0(F)$, is maintained. The solutions $Pr^*(X)$ to eqn (A1) are entered for each limit distribution of step 1. For example, the highest proportion of teen-aged drivers for which a $Pr^*(F)$ is reported in the first procedure suggests coefficients for an eqn (4) to be solved for each other margin (sex, BAL, seatbelt use, vehicle type, and number of occupants).

Because each variation in a first margin corresponds to coefficients of eqn (A2) for each of five other margins, there are five solution sets. Each set has dimension $N_i - 1$ where $N_i$ is the number of categories in the $i$th margin. In each of these solution sets is a point that is closest to the NRBS value for the corresponding margin. These points are approximated by a grid search procedure with increments of .02 along each coordinate. The solution reported for each second margin is the point in the eqn (A2) solution set found to be closest to the NRBS probabilities. The Mahalanobis distance function [Eqn (5)] is used to determine distance to this closest point. It may happen that the pair of margins $P_i^*$ and $P_j^*$ that satisfy eqn (A2) cause distortion in the fitting procedure and lead to a Pr(F) in eqn (2) that is different from the starting $Pr_0(F)$. This distortion becomes more extreme as marginal distributions depart from NRBS values. Limits found in step one are modified as required, so that the Pr(F) solving eqn (3) in step 2 was within 1% of $Pr_0(F)$. All possible ordered pairs of driver characteristics are so evaluated.

## APPENDIX B: CHANGES TO TWO MARGINAL DISTRIBUTIONS THAT CORRESPOND TO $Pr_0(F)$

Numbers reported in ( ) represent the following:

$$
\begin{aligned}
\text{Age} &= Pr(\text{under 20 years}), Pr(20 \text{ to } 29), Pr(30 \text{ to } 39), Pr(40 \text{ and over}) \\
\text{Vehicle} &= Pr(\text{small car}), Pr(\text{medium size}), Pr(\text{large car}), Pr(\text{truck or van}) \\
\text{Occupants} &= Pr(\text{one occupant}), Pr(2 \text{ occupants}), Pr(3 \text{ or more}) \\
\text{Sex} &= Pr(\text{male driver}), Pr(\text{female}) \\
\text{Belt} &= Pr(\text{belt used}), Pr(\text{not used}) \\
\text{BAL} &= Pr(\text{BAL} = 0), Pr(\text{BAL between 0 and .02 mg/dL}), \text{mean BAL for BAL over 0.}
\end{aligned}
$$

A: *Age margin from* (.213 .369 .182 .236) *to*
   A-1 *Reduce proportion teen-aged drivers*: (.213 .408 .201 .260) for Pr(F) −1.98%.
   Then recover $Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .345 \quad .380 \quad .095) \\
\text{Occupants} &= (.359 \quad .360 \quad .281) \\
\text{Sex} &= (.876 \quad .124) \\
\text{Belt} &= (.300 \quad .700) \\
\text{BAL} &= (.526 \quad .240 \quad .0383)
\end{aligned}
$$

*Increase proportion teen-aged drivers*: (.290 .333 .165 .212) for Pr(F) + 2.05%.
Then recover $Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .340 \quad .402 \quad .078) \\
\text{Occupants} &= (.331 \quad .360 \quad .309) \\
\text{Sex} &= (.736 \quad .264) \\
\text{Belt} &= (.316 \quad .684) \\
\text{BAL} &= (.526 \quad .240 \quad .0369)
\end{aligned}
$$

A-2 *Reduce proportion 20–29 year olds*: (.263 .221 .225 .291) for Pr(F) −4.18%.
Then recover $Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .360 \quad .368 \quad .092) \\
\text{Occupants} &= (.383 \quad .360 \quad .257) \\
\text{Sex} &= (.981 \quad .019) \\
\text{Belt} &= (.266 \quad .734) \\
\text{BAL} &= (.526 \quad .240 \quad .0385)
\end{aligned}
$$

*Increase proportion 20–29 year olds*: (.172 .490 .148 .190) for Pr(F) +3.60%.
Then recover $Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .340 \quad .406 \quad .074) \\
\text{Occupants} &= (.318 \quad .360 \quad .322) \\
\text{Sex} &= (.699 \quad .301) \\
\text{Belt} &= (.341 \quad .659) \\
\text{BAL} &= (.526 \quad .240 \quad .0368)
\end{aligned}
$$

A-3 *Reduce proportion 30–39 year olds*: (.244 .423 .062 .270) for $\Pr(F)$ $-1.58\%$.
Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .342 \quad .400 \quad .078) \\
\text{Occupants} &= (.335 \quad .360 \quad .305) \\
\text{Sex} &= (.757 \quad .243) \\
\text{Belt} &= (.327 \quad .673) \\
\text{BAL} &= (.526 \quad .240 \quad .0366)
\end{aligned}
$$

*Increase proportion 30–39 year olds*: (.178 .308 .317 .197) for $\Pr(F)$ $-1.55\%$.

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .348 \quad .380 \quad .092) \\
\text{Occupants} &= (.355 \quad .360 \quad .285) \\
\text{Sex} &= (.870 \quad .130) \\
\text{Belt} &= (.287 \quad .713) \\
\text{BAL} &= (.526 \quad .240 \quad .0386)
\end{aligned}
$$

A-4 *Reduce proportion 40 and older*: (.241 .418 .207 .134) for $\Pr(F)$ $+5.28\%$.
Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.179 \quad .320 \quad .420 \quad .081) \\
\text{Occupants} &= (.303 \quad .360 \quad .337) \\
\text{Sex} &= (.637 \quad .363) \\
\text{Belt} &= (.339 \quad .660) \\
\text{BAL} &= (.526 \quad .240 \quad .0366)
\end{aligned}
$$

*Increase proportion 40 and older*: (.244 .423 .062 .270) for $\Pr(F)$ $-5.06\%$.
Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Vehicle} &= (.180 \quad .340 \quad .372 \quad .108) \\
\text{Occupants} &= (.387 \quad .360 \quad .253) \\
\text{Sex} &= (.994 \quad .006) \\
\text{Belt} &= (.274 \quad .726) \\
\text{BAL} &= (.526 \quad .240 \quad .0386)
\end{aligned}
$$

B. *Vehicle margin from* (.178 .342 .393 .087) *to*
  B-1 *Reduce proportion small cars*: (.004 .414 .476 .106) for $\Pr(F)$ $+5.11\%$.
  Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Age} &= (.200 \quad .299 \quad .180 \quad .321) \\
\text{Occupants} &= (.307 \quad .333 \quad .370) \\
\text{Sex} &= (.697 \quad .303) \\
\text{Belt} &= (.345 \quad .665) \\
\text{BAL} &= (.526 \quad .240 \quad .0369)
\end{aligned}
$$

*Increase proportion small cars*: (.294 .394 .337 .075) for $\Pr(F)$ $-3.26\%$.
Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Age} &= (.220 \quad .420 \quad .180 \quad .180) \\
\text{Occupants} &= (.374 \quad .360 \quad .266) \\
\text{Sex} &= (.981 \quad .264) \\
\text{Belt} &= (.282 \quad .718) \\
\text{BAL} &= (.526 \quad .240 \quad .0380)
\end{aligned}
$$

B-2 *Reduce proportion medium cars*: (.189 .300 .418 .193) for $\Pr(F)$ $-4.42\%$.
Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Age} &= (.220 \quad .446 \quad .180 \quad .154) \\
\text{Occupants} &= (.384 \quad .360 \quad .256) \\
\text{Sex} &= (.975 \quad .025) \\
\text{Belt} &= (.272 \quad .728) \\
\text{BAL} &= (.526 \quad .240 \quad .0382)
\end{aligned}
$$

*Increase proportion medium cars*: (.076 .720 .167 .037) for $\Pr(F)$ $+40.04\%$.
Then recover $\Pr_0(F)$ when

$$
\begin{aligned}
\text{Age} &= (.005 \quad .160 \quad .180 \quad .655) \\
\text{Occupants} &= (.106 \quad .360 \quad .533) \\
\text{Sex} &= (.003 \quad .997) \\
\text{Belt} &= (.528 \quad .472) \\
\text{BAL} &= (.526 \quad .240 \quad .0329)
\end{aligned}
$$

B-3 *Reduce proportion large cars*: (.235 .453 .196 .116) for Pr(F) +26.28%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rllll}
\text{Age} = & (.007 & .220 & .180 & .523) \\
\text{Occupants} = & (.176 & .360 & .464) \\
\text{Sex} = & (.093 & .907) \\
\text{Belt} = & (.468 & .532) \\
\text{BAL} = & (.526 & .240 & .0344)
\end{array}
$$

*Increase proportion small cars*: (.166 .320 .432 .082) for Pr(F) −5.15%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rllll}
\text{Age} = & (.217 & .460 & .180 & .143) \\
\text{Occupants} = & (.391 & .360 & .249) \\
\text{Sex} = & (.997 & .003) \\
\text{Belt} = & (.266 & .734) \\
\text{BAL} = & (.526 & .240 & .0383)
\end{array}
$$

B-4 *Reduce proportion trucks and vans*: (.183 .554 .406 .057) for Pr(F) −4.52%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rllll}
\text{Age} = & (.219 & .440 & .180 & .161) \\
\text{Occupants} = & (.386 & .360 & .254) \\
\text{Sex} = & (.988 & .012) \\
\text{Belt} = & (.272 & .728) \\
\text{BAL} = & (.526 & .240 & .0382)
\end{array}
$$

*Increase proportion trucks and vans*: (.152 .292 .336 .220) for Pr(F) + 20.39%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rllll}
\text{Age} = & (.200 & .008 & .180 & .612) \\
\text{Occupants} = & (.211 & .360 & .429) \\
\text{Sex} = & (.988 & .012) \\
\text{Belt} = & (.272 & .728) \\
\text{BAL} = & (.526 & .240 & .0382)
\end{array}
$$

*Increase proportion trucks and vans*: (.151 .292 .336 .220) for Pr(F) + 20.39%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rllll}
\text{Age} = & (.200 & .008 & .180 & .612) \\
\text{Occupants} = & (.211 & .360 & .429) \\
\text{Sex} = & (.377 & .620) \\
\text{Belt} = & (.437 & .536) \\
\text{BAL} = & (.526 & .240 & .0351)
\end{array}
$$

C. *Number occupants margin from* (.347 .358 .295) *to*
   C-1 *Reduce proportion 1 occupant*: (.305 .381 .313) for Pr(F) −4.46%.
   Then recover Pr(F) when

$$
\begin{array}{rlllll}
\text{Age} = & (.280 & .374 & .180 & .166) \\
\text{Vehicle} = & (.180 & .360 & .364 & .096) \\
\text{Sex} = & (.978 & .022) \\
\text{Belt} = & (.272 & .728) \\
\text{BAL} = & (.526 & .240 & .0384)
\end{array}
$$

*Increase proportion 1 occupant*: (.645 .195 .160) for Pr(F) +32.68%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rlllll}
\text{Age} = & (.020 & .008 & .020 & .951) \\
\text{Vehicle} = & (.180 & .300 & .514 & .006) \\
\text{Sex} = & (.110 & .890) \\
\text{Belt} = & (.496 & .504) \\
\text{BAL} = & (.526 & .240 & .0330)
\end{array}
$$

C-2 *Reduce proportion 2 occupants*: (.305 .381 .303) for Pr(F) + 15.22%.
Then recover $Pr_0(F)$ when

$$
\begin{array}{rlllll}
\text{Age} = & (.212 & .080 & .180 & .528) \\
\text{Vehicle} = & (.180 & .340 & .451 & .027) \\
\text{Sex} = & (.388 & .612) \\
\text{Belt} = & (.408 & .592) \\
\text{BAL} = & (.526 & .240 & .0357)
\end{array}
$$

*Increase proportion* 2 *occupants:* (.289 .465 .246) for Pr(F) −4.50%.
Then recover Pr$_0$(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.280 & .371 & .180 & .169) \\
\text{Vehicle} & = (.180 & .360 & .364 & .096) \\
\text{Sex} & = (.980 & .020) \\
\text{Belt} & = (.272 & .728) \\
\text{BAL} & = (.526 & .240 & .0382)
\end{array}
$$

C-3 Reduce proportion 3 *or more:* (.489 .505 .006) for Pr(F) +20.75%.
Then recover Pr(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.200 & .040 & .181 & .578) \\
\text{Vehicle} & = (.180 & .340 & .468 & .012) \\
\text{Sex} & = (.285 & .715) \\
\text{Belt} & = (.440 & .560) \\
\text{BAL} & = (.526 & .240 & .0341)
\end{array}
$$

*Increase proportion* 3 *or more:* (.318 .328 .354) for Pr(F) −4.15%.
Then recover Pr$_0$(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.216 & .440 & .180 & .164) \\
\text{Vehicle} & = (.180 & .360 & .366 & .094) \\
\text{Sex} & = (.966 & .034) \\
\text{Belt} & = (.274 & .756) \\
\text{BAL} & = (.526 & .240 & .0384)
\end{array}
$$

D. *Sex of driver margins from* (.808 .192) *to*
*Reduce proportion males:* (.700 .300) for Pr(F) −3.01%.
Then recover Pr$_0$(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.218 & .410 & .180 & .182) \\
\text{Vehicle} & = (.180 & .340 & .380 & .100) \\
\text{Occupants} & = (.382 & .360 & .258) \\
\text{Belt} & = (.284 & .716) \\
\text{BAL} & = (.526 & .240 & .0374)
\end{array}
$$

*Increase proportion males:* (.908 .092) for Pr(F) +2.98%.
Then recover Pr$_0$(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.200 & .335 & .180 & .285) \\
\text{Vehicle} & = (.180 & .340 & .403 & .077) \\
\text{Occupants} & = (.316 & .360 & .324) \\
\text{Belt} & = (.329 & .671) \\
\text{BAL} & = (.526 & .240 & .0377)
\end{array}
$$

E. *Seat belt use margins from* (.308 .692) *to*
*Reduce proportion using belt:* (.134 .866) for Pr(F) +22.81%.
Then recover Pr$_0$(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.206 & .060 & .180 & .553) \\
\text{Vehicle} & = (.180 & .340 & .477 & .003) \\
\text{Occupants} & = (.194 & .360 & .446) \\
\text{Sex} & = (.157 & .843) \\
\text{BAL} & = (.526 & .240 & .0359)
\end{array}
$$

*Increase proportion using belt:* (.348 .652) for Pr(F) −5.10%.
Then recover Pr$_0$(F) when

$$
\begin{array}{lllll}
\text{Age} & = (.216 & .460 & .180 & .144) \\
\text{Vehicle} & = (.180 & .360 & .361 & .099) \\
\text{Occupants} & = (.391 & .360 & .249) \\
\text{Sex} & = (.994 & .006) \\
\text{BAL} & = (.526 & .240 & .0380)
\end{array}
$$