U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

People Saving People
http://www.nhtsa.dot.gov

DOT HS 809 403 **January 2002**
*(Revised October, 2002)**

Technical Report

Transitioning to Multiple Imputation – A New Method to Impute Missing Blood Alcohol Concentration (BAC) values in FARS

Published By:

NCSA
**National Center for Statistics and Analysis
Research and Development**

** This report was revised [June, 2002] from its original version [January, 2002] to reflect corrected crash-level alcohol estimates from Multiple Imputation for 1998-2000 and again revised [October, 2002] to reflect updated programs to analyze the imputed data.*

Technical Report Documentation Page

1. Report No. DOT HS 809 403	2. Government Accession No.	3. Recipients's Catalog No.
4. Title and Subtitle Transitioning to Multiple Imputation – A New Method to Estimate Missing Blood Alcohol Concentration (BAC) values in FARS		5. Report Date January, 2002
		6. Performing Organization Code NRD-31
7. Author(s) Rajesh Subramanian		8. Performing Organization Report No.
9. Performing Organization Name and Address Mathematical Analysis Division, National Center for Statistics and Analysis National Highway Traffic Safety Administration U.S. Department of Transportation NRD-31, 400 Seventh Street, S.W. Washington, DC 20590		10. Work Unit No. (TRAIS)n code
		11. Contract of Grant No.
12. Sponsoring Agency Name and Address Mathematical Analysis Division, National Center for Statistics and Analysis National Highway Traffic Safety Administration U.S. Department of Transportation NRD-31, 400 Seventh Street, S.W. Washington, DC 20590		13. Type of Report and Period Covered NHTSA Technical Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The National Center for Statistics and Analysis (NCSA) of the National Highway Traffic Safety Administration (NHTSA) has undertaken several approaches to remedy the problem of missing blood alcohol test results in the Fatality Analysis Reporting System (FARS). The current approach employs a linear discriminant model that estimates the probability that a driver or nonoccupant has a BAC in grams per deciliter (g/dl) of 0, .01 to .09 or .10 and greater. Estimates are generated only for drivers and nonoccupants (pedestrians, pedalcyclists) for whom alcohol test results were not reported.

Beginning with the 2001 data, NHTSA will transition to Multiple Imputation, a new method to estimate missing BAC in FARS. The publications for the 2001 data will reflect the estimates of alcohol involvement generated using Multiple Imputation. The new methodology improves on the current model by imputing specific values of BAC across the full range of possible values rather than estimating probabilities. Imputing ten values of BAC for each missing value will permit the estimation of valid statistics such as variances, measures of central tendency, confidence intervals and standard deviations.

17. Key Words Alcohol, BAC, Multiple Imputation, Fatal Crashes		18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 36	22. Price

Form DOT F1700.7 (8-72) **Reproduction of completed page authorized**

Table of Contents

1. Introduction

The National Highway Traffic Safety Administration (NHTSA) through the National Center for Statistics and Analysis (NCSA) has undertaken several approaches to remedy the problem of missing blood alcohol test results in the Fatality Analysis Reporting System (FARS). The current approach employs a linear discriminant model that estimates the probability that a driver or nonoccupant has a BAC in grams per deciliter (g/dl) of 0, .01 to .09 or .10 and greater. Estimates are generated only for drivers and nonoccupants (pedestrians, pedalcyclists) for whom alcohol test results were not reported.

Beginning with the 2001 data, NCSA will transition to Multiple Imputation, a new method to estimate missing BAC in FARS. The Multiple Imputation procedure in FARS has been implemented and validated as outlined in an earlier NHTSA Technical Report (Rubin et al., [8]). The new methodology improves on the current model by imputing specific values of BAC across the full range of possible values rather than estimating probabilities. Imputing ten values of BAC for each missing value will permit the estimation of valid statistics such as variances, measures of central tendency, confidence intervals and standard deviations. As a result, researchers will be able to draw inferences and test hypotheses about BAC as a factor in fatal traffic crashes. The estimation of discrete values also facilitates analysis by nonstandard boundaries of alcohol involvement (e.g., .08+).

This report outlines the Multiple Imputation methodology and explains the reasons for and the advantages of shifting from the old methodology of estimating missing BAC. It also documents the results from both methodologies for certain categories from the 1982 to the 2000 data years. The estimates are based on the Final FARS data for 1982 through 1999 and the Annual Report File (ARF) for 2000. State-by-state estimates of alcohol involvement from the two methods are also presented in this report for the 1999 data year.

2. Reasons for changing the Imputation methodology

Multiple Imputation is a state of the art methodology - Imputation of missing data points is a constantly evolving field of statistics. When the discriminant analysis methodology was adopted in 1986, NHTSA recognized the need to constantly evaluate alternative methods of imputation to improve the quality and usefulness of our data. The current method permits the calculation of probabilities that a missing value falls within one of the three ranges. The inability to impute discrete values with this method greatly limits analysis of the effects of alcohol involvement on traffic crashes and fatalities. The quest for a technique that will provide the most statistically robust estimates of discrete values has led to Multiple Imputation. Multiple Imputation is the state-of-the-art imputation methodology that will permit computation of standard errors of estimates and confidence intervals, thus enabling researchers to test hypotheses.

NHTSA assembled a panel of leading experts in different forms of imputation who evaluated the available methodologies and their applicability to FARS data. The panel recommended Multiple

Imputation unanimously. The work of this panel was summarized in a NHTSA document (de Wolf, [1]) that focuses on the major recommendations of the panelists. The panel members outlined their recommendations in papers submitted to NHTSA (Huberty, C.J. et. al. [2], Kalton, G. et. al. [3], Little, R. [5], Rubin, D.B. [7]). Multiple Imputation has already been implemented to solve nonresponse problems in various Government and Private data systems. It is a procedure that has been peer-reviewed and published in leading professional journals.

Changing legislative needs – At the federal level, a BAC of .08 has been established as the legal definition of intoxication. Through the use of fiscal incentives, the states are being encouraged to adopt this threshold as a standard. Using the current Discriminant Analysis methodology, NCSA has no way to estimate alcohol involvement at .08 and above without significantly altering the method. Even with modification, the method of estimation might not be consistent with the process used in the earlier years, which will affect the reporting of trends in alcohol involvement.

Multiple Imputation permits statistics along any BAC standard as it imputes BAC along the entire scale of plausible BAC values. A historic revision of estimates up to the 1982 data also ensures the consistency of estimates of trends.

3. Advantages of the new method

Multiple Imputation provides significant advantages to NCSA in its efforts to address changing research and reporting needs. The imputed values are actual values of BAC along the entire plausible range, and can therefore be combined to provide estimates for any defined category of alcohol involvement. Analysts can use the imputed values to include BAC in models and to prepare estimates of error attributable to the imputation process. These procedures were not possible with the Discriminant Analysis method.

The multiple imputation procedure replaces each missing BAC with ten simulated values. The ten imputations, together with the non-missing BAC values, produce ten apparently complete versions of BAC, each of which may be analyzed by standard complete-data techniques. Results from analyzing the ten versions will vary somewhat. This variation is used to estimate the extra uncertainty in statistical summaries that is attributable to missing data. The ten sets of answers are combined with simple computational macros implementing rules given by Rubin [6]. Combining the ten answers according to these special rules produces statistical inferences that are valid (i.e., estimation of parameters that are consistent, nominal 95% confidence intervals are in fact 95% confidence intervals etc.) under quite general conditions. This offers an advantage over single imputation which, even if done properly to allow consistent estimation of parameters, uniformly underestimates variability because one imputed value cannot possibly represent uncertainty.

The Discriminant Analysis method estimates the probabilities that the actual value falls within each of the three broad BAC classes. Although limiting in ways as discussed in Section 2, this method is preferable to single imputation. Nevertheless, the estimated probabilities cannot be used for many complete-data analyses, especially those pertaining to other categorizations of BAC, such as the .08+ category. Moreover, they do not reflect any uncertainty about the estimated parameters in the discriminant models used to compute the probabilities.

4. Analyzing the Multiply-imputed Datasets

When assessing the extent of alcohol involvement in traffic crashes, the quantity of interest is usually the proportion of a population that shows the involvement of alcohol (e.g., percent of drivers killed that were intoxicated, percent of fatally injured nonoccupants, etc). This proportion is the percentage of the standard population of the stratum of interest that has alcohol involvement. Alcohol involvement is determined jointly from the known set of alcohol test results as well as the imputed values for unknown BAC. Under multiple imputation, each missing BAC value is replaced by ten imputed values. In order to estimate population proportions, the results (proportions) from each of the ten sets of values have to be combined by standard computational macros.

Rubin's method of scalar estimands (Rubin, [6]) is used to estimate quantities of interest. Let Q be a one-dimensional quantity of interest – a *proportion* of crashes or persons that showed a positive alcohol test result in a universe of crashes or people or a *coefficient* from a linear or logistic regression model. The goal is to find a confidence interval or test a hypothesis about Q. Let ***Y*** denote the data from FARS that are necessary to estimate ***Q***. ***Y*** is partitioned into observed and missing parts,

$$Y = (Y_{obs}, Y_{mis})$$

where Y_{obs} is known and Y_{mis} is unknown and has been multiply-imputed. Let $\hat{Q}$ be the complete-data point estimate for Q, the estimate to be used if no data were missing. Let U be the variance estimate associated with $\hat{Q}$, so that $\sqrt{U}$ is the complete-data standard error. As U and $\hat{Q}$ are both functions of $Y = (Y_{obs}, Y_{mis})$, they may be rewritten as $\hat{Q}(Y_{obs}, Y_{mis})$ and $U(Y_{obs}, Y_{mis})$, respectively. Multiple Imputation inference assumes that the complete data problem is sufficiently regular and sample size sufficiently large for the asymptotic normal approximation

$$U^{-1/2}(Q - \hat{Q}) \sim N(0,1)$$

to work well. With m imputations, m different versions of $\hat{Q}$ and U can be calculated.
Let

$$\hat{Q}^{(t)} = \hat{Q}(Y_{obs}, Y_{mis}^{(t)})$$

and

$$U^{(t)} = U(Y_{obs}, Y_{mis}^{(t)})$$

be the point and variance estimates using the t-*th* set of imputed data, t=1,2,...,10. The multiple imputation point-estimate for Q is simply the average of the complete-data point estimates.

$$\overline{Q} = \frac{1}{10}\sum_{i=1}^{10} \hat{Q}^{(t)}$$

$\overline{Q}$ is the final quantity of interest, for example, the proportion of drivers involved in fatal crashes whose BAC was .01 or above.

The variance estimate associated with $\overline{Q}$ has two components. The *within-imputation* variance is the average of the complete-data variance estimates,

$$\overline{U} = \frac{1}{10}\sum_{t=1}^{10} U^{(t)}$$

and the *between-imputation* variance is the variance of the complete-data point estimates,

$$B = \frac{1}{9}\sum_{t=1}^{10} (\hat{Q}^{(t)} - \overline{Q})^2$$

The *total-variance* is defined as

$$T = \overline{U} + (1 + m^{-1})B$$

5. Comparison of estimates from the two methods

Under Multiple Imputation, the estimated rates of alcohol involvement were generally higher than those under Discriminant Analysis. Positive differences of up to about 2 percent in the rate of alcohol involvement appeared consistently across most of the broader categories of vehicle classes and demographic subgroups, and across classifications of crashes by time of day and day of week. Differences in rates across subgroups, and trends in rates across time, were quite similar under the two methods.

The discrepancy between the new and old imputation methods can be traced to the fact that the new model is a General Linear Location Model (GLOM) which is a two-stage model[1] while the

[1] The two-stage model consists of a first stage conventional loglinear model to determine BAC as a binary indicator, i.e., if BAC=0 or BAC>0. The second stage of the GLOM is a conventional linear regression model used to predict the actual level of BAC given that BAC>0 (Rubin et al., 1998).

old model was a discriminant model. To the uninformed observer, it may seem disturbing that changing the form of the missing data model would have so much impact on the final estimates. However, given the high rates of missing information (>50%) about BAC, the results should be sensitive to model specification. This underscores the need to provide meaningful estimates of uncertainty (e.g. standard errors) for statistics related to BAC. The new method addresses this issue and can now provide standard errors of estimates of alcohol involvement.

Also, the old imputation method is based on a linear discriminant model that distinguishes among the three classes of BAC. The new model, however, models BAC in two stages: a logit model for distinguishing BAC=0 and BAC>0, and a regression model for predicting BAC given that BAC>0. In many situations, discriminant and logit models tend to produce similar estimates for the classification probabilities. On occasion, however, the probabilities can be different because of a critical assumption that the discriminant model makes. The assumption is that the normal distributions for the three classifications share a common covariance matrix. If the group means are far apart, then the classification probabilities from the linear discriminant model can be substantially less efficient than those of the logit model.

Tables 1 through 4 present a comparison of the BAC estimates from the multiple imputation method to those of the discriminant analysis (Klein, [4]) model under various categories traditionally reported and released by NHTSA in its annual fact sheets. The estimates from multiple imputation are printed in each cell accompanied by the corresponding estimates from the discriminant method in parenthesis. The values in each cell are the percentage of all crashes or drivers in that category for two levels of BAC, namely, .01 and greater (.01+)[2] and .10 and greater (.10+). For example, an entry of 36.2 for a given year for drivers killed in the .01+ category means that for that particular year, 36.2 percent of all the drivers killed in fatal crashes had a BAC of .01 or greater.

Table 1 presents the overall rates of alcohol involvement and intoxication in fatal crashes as estimated by the two methods. Alcohol is said to be involved in a fatal traffic crash if either the driver of any vehicle or a nonoccupant (pedestrian or pedalcyclist) involved in the crash has a BAC of .01 or greater. A driver or nonoccupant is deemed "intoxicated" if the BAC is .10 or greater. However, according to new legislation, the legal threshold for intoxication is .08. In the tables, figures in parentheses represent the values obtained using Discriminant Analysis.

[2] A BAC level of 0.01 implies that the alcohol content is 0.01 grams/deciliter. Because BAC is reported to two decimal places, the 0.01+ category includes all positive values of BAC.

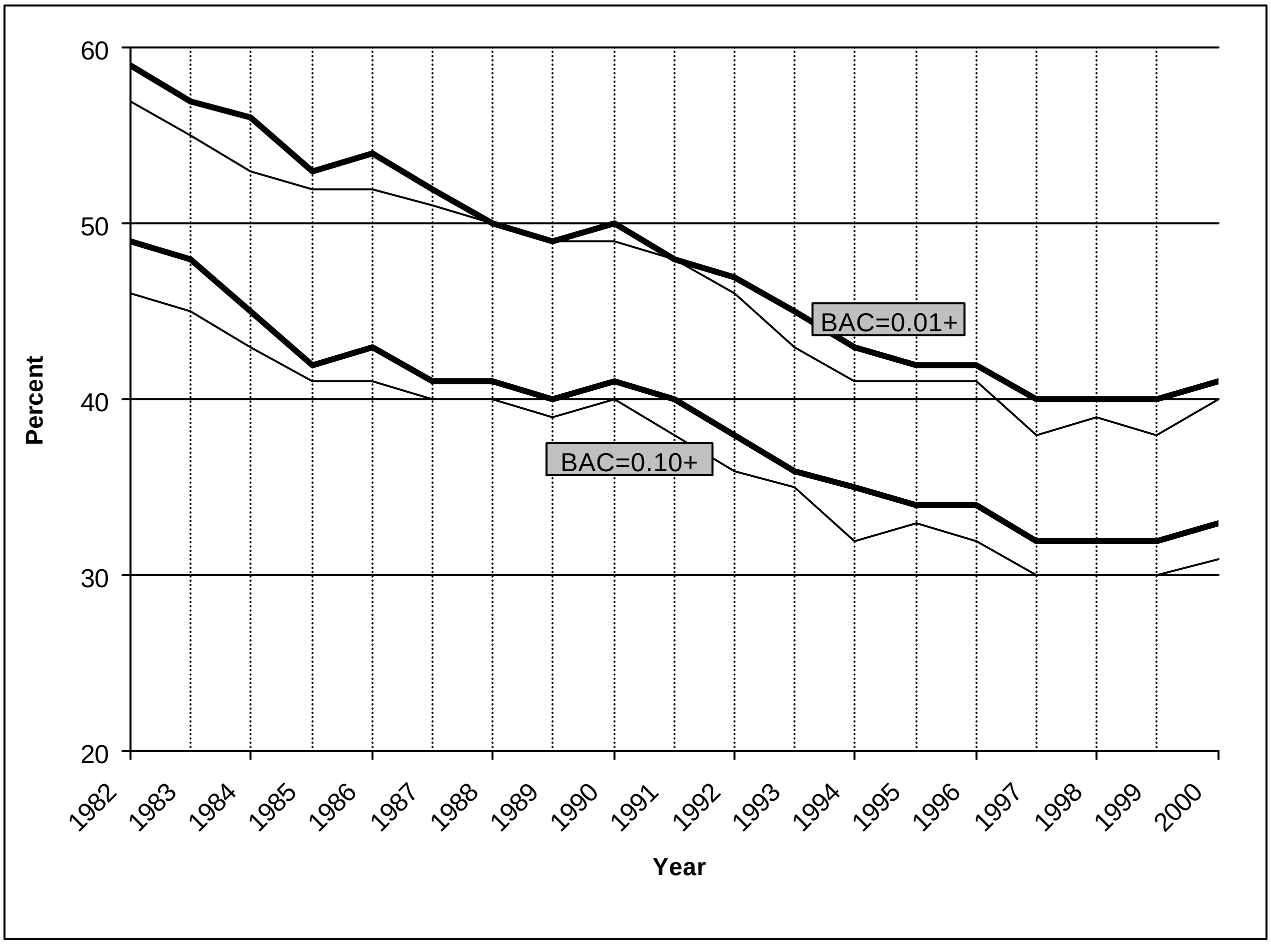
Table 1 : Alcohol Involvement (Percentage) in Fatal Crashes and Fatalities by Crash BAC and Imputation Methodology, FARS 1982-2000*

Year	Fatal Crashes		Fatalities	
	BAC=.01+	BAC=.10+	BAC=.01+	BAC=.10+
1982	59 (57)	49 (46)	60 (57)	49 (46)
1983	57 (55)	48 (45)	58 (56)	48 (45)
1984	56 (53)	45 (43)	56 (54)	46 (43)
1985	53 (52)	42 (41)	53 (52)	43 (41)
1986	54 (52)	43 (41)	54 (52)	43 (41)
1987	52 (51)	41 (40)	52 (51)	41 (40)
1988	50 (50)	41 (40)	51 (50)	41 (40)
1989	49 (49)	40 (39)	49 (49)	40 (39)
1990	50 (49)	41 (40)	51 (50)	41 (40)
1991	48 (48)	40 (38)	49 (48)	40 (38)
1992	47 (46)	38 (36)	47 (45)	38 (36)
1993	45 (43)	36 (35)	45 (44)	36 (35)
1994	43 (41)	35 (32)	43 (41)	34 (32)
1995	42 (41)	34 (33)	42 (41)	34 (32)
1996	42 (41)	34 (32)	42 (41)	34 (32)
1997	40 (38)	32 (30)	40 (39)	32 (30)
1998	40 (39)	32 (30)	40 (39)	32 (30)
1999	40 (38)	32 (30)	40 (38)	32 (30)
2000	41 (40)	33 (31)	41 (40)	33 (31)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
*(Values in parentheses represent estimates from the **old** imputation methodology)*

Figure 1 illustrates the differences in the trend of alcohol involvement in fatal crashes from 1982 to 2000. The trend of alcohol involvement follows the same pattern for the estimates from both the methods. The discrepancy may be higher in the earlier years due to the high degree of missing information in the earlier years. Using the estimates based on Multiple Imputation, trend lines can be generated for any level of alcohol along the range of plausible values (0 to .94).

Figure 1 : Alcohol Involvement in Fatal Crashes by Crash BAC and Imputation Methodology, FARS 1982-2000



95% Confidence Intervals and Standard Errors

The ten sets of imputations are combined with simple computational macros implementing rules given by Rubin [6]. Combining the ten answers according to these special rules produces statistical inferences that are valid (i.e., estimation of parameters that are consistent, nominal 95% confidence intervals are in fact 95% confidence intervals etc.) under quite general conditions. The total variance T estimated above is used in evaluating the 95% confidence intervals. The inferences are based on the assumption that

$$T^{-1/2}(Q-\overline{Q}) \sim t_n$$

where the degrees of freedom < are given by

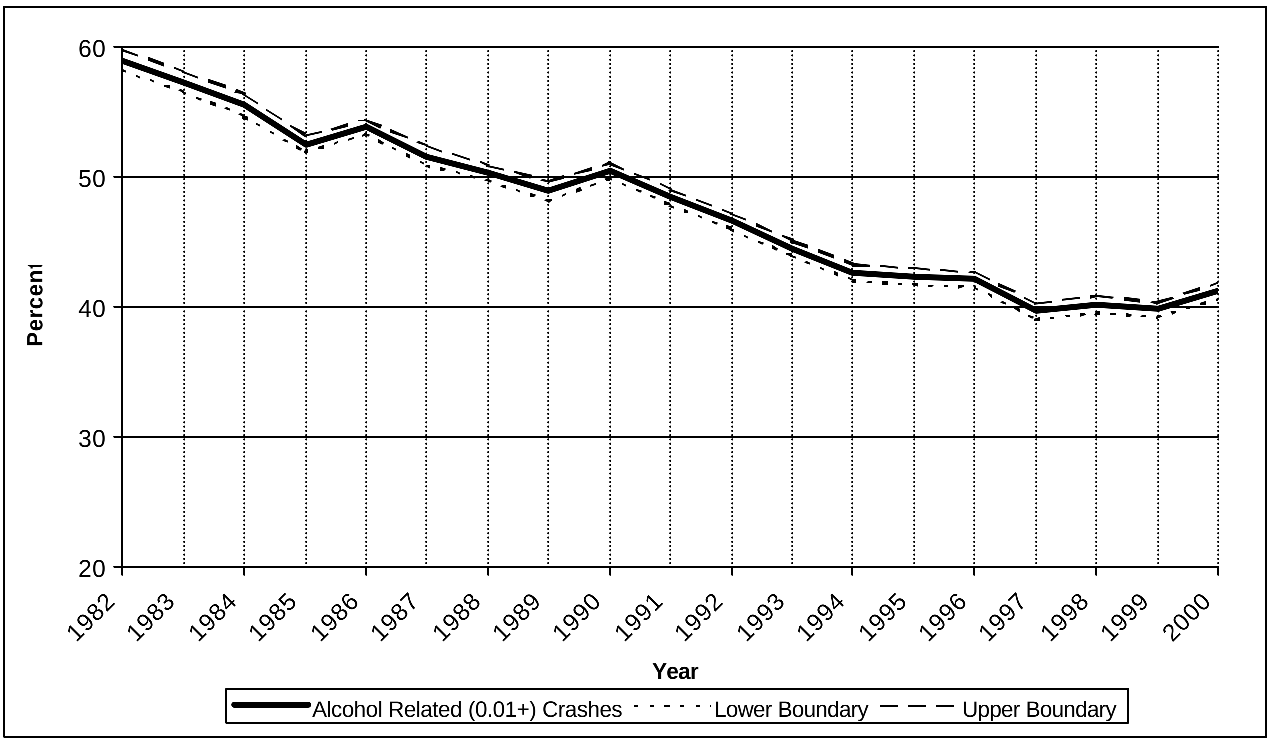
$$n = (m-1)[1+\frac{\overline{U}}{(1+m^{-1})B}]^2$$

Thus a 100(1-")% confidence-interval estimate for the estimate is given by

$$\overline{Q} \pm t_{n,1-a/2}\sqrt{T}$$

Figure 2 depicts the 95% Confidence Interval band of the estimated percentage of alcohol involvement (.01+) in fatal crashes. The upper and lower bounds are estimated according to the SAS procedures listed in Appendix B.

Figure 2 : 95% Confidence Interval of the estimated percentage of Alcohol Involvement in Fatal Crashes by Crash BAC, FARS 1982-2000



The standard error of the estimate is the square-root of the total variance T, as estimated by

$$Standard\ Error = \sqrt{\overline{U} + (1 + m^{-1})B}$$

Table 3 shows the Standard Error of the estimated alcohol involvement in fatal crashes by Crash BAC.

Table 3 : Standard Error of the Estimate of Alcohol Involvement (Percentage) in Fatal Crashes by Crash BAC, FARS 1982-2000*

Year	Estimate	Standard Error
1982	59	0.39435745
1983	57	0.38229963
1984	56	0.41550553
1985	53	0.30986141
1986	54	0.26917724
1987	52	0.36467598
1988	50	0.29723237
1989	49	0.36094569
1990	50	0.29752044
1991	48	0.31703266
1992	47	0.30584581
1993	45	0.29541456
1994	43	0.30842406
1995	42	0.31074702
1996	42	0.29497051
1997	40	0.31334548
1998	40	0.32099755
1999	40	0.27862571
2000	41	0.32049002

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*

Table 2 shows the trend of fatally injured persons in crashes where at least one driver or nonoccupant (pedestrian or pedalcyclist) was intoxicated. On average, about 70 percent of people killed in such crashes were themselves intoxicated. The remaining 30 percent were passengers, nonintoxicated drivers, or nonintoxicated nonoccupants.

Table 2: Types of Fatalities in Fatal Crashes Involving at Least One Intoxicated Driver or Nonoccupant Expressed as a Percent of Such fatalities, FARS 1982-2000*

Year	Intoxicated Drivers	Nonintoxicated Drivers	Passengers	Intoxicated Nonoccupants	Nonintoxicated Nonoccupants
1982	52 (53)	7 (7)	22 (21)	13 (13)	5 (6)
1983	52 (53)	7 (7)	23 (22)	12 (12)	5 (5)
1984	53 (54)	8 (7)	21 (21)	12 (13)	5 (6)
1985	54 (54)	8 (7)	22 (21)	12 (13)	5 (5)
1986	54 (54)	8 (7)	22 (22)	12 (12)	5 (5)
1987	54 (55)	8 (7)	22 (21)	12 (12)	5 (5)
1988	54 (55)	8 (7)	22 (21)	12 (12)	4 (5)
1989	55 (55)	7 (7)	22 (21)	12 (12)	4 (5)
1990	53 (55)	7 (6)	22 (21)	12 (12)	5 (5)
1991	54 (55)	7 (7)	22 (22)	12 (13)	4 (4)
1992	54 (54)	7 (7)	22 (21)	13 (13)	4 (4)
1993	54 (54)	7 (7)	21 (21)	14 (14)	4 (4)
1994	53 (56)	8 (7)	22 (20)	13 (13)	5 (4)
1995	55 (56)	7 (6)	20 (20)	13 (14)	4 (4)
1996	54 (55)	7 (6)	22 (21)	13 (14)	4 (4)
1997	55 (55)	7 (6)	21 (21)	13 (13)	4 (4)
1998	56 (56)	6 (6)	21 (20)	14 (14)	4 (4)
1999	56 (57)	7 (6)	21 (20)	13 (14)	3 (4)
2000	55 (57)	7 (6)	22 (21)	12 (12)	3 (4)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
(values in parentheses represent estimates from the ***old*** *imputation methodology)*

Table 3 compares the results obtained by the two methods for drivers only. The table shows BAC levels, by year, for drivers who were either involved in a fatal crash or killed in a fatal crash. As in the previous table, the figures in parentheses represent the values obtained by the current method of imputation. Alcohol involvement for drivers involved in fatal crashes, who themselves may be fatally injured or have survived, ranged from a high of about 41 percent in 1982 to a low of about 24 percent in 1999. This trend was also reflected when looking at just male or female drivers involved in fatal crashes.

Table 3 : Alcohol Involvement for Drivers Involved in Fatal Crashes by Sex, FARS 1982-2000*

Year	Male		Female		Total	
	.01+	.10+	.01+	.10+	.01+	.10+
1982	44 (42)	35 (32)	27 (26)	20 (19)	41 (39)	32 (30)
1983	43 (40)	35 (31)	25 (25)	20 (18)	39 (38)	32 (29)
1984	41 (39)	32 (30)	25 (24)	18 (17)	38 (36)	29 (27)
1985	38 (37)	30 (28)	22 (22)	16 (15)	35 (34)	27 (26)
1986	40 (38)	30 (29)	22 (21)	16 (15)	36 (34)	27 (26)
1987	37 (36)	29 (28)	21 (21)	16 (15)	34 (33)	26 (25)
1988	37 (36)	29 (28)	20 (20)	15 (15)	33 (33)	26 (25)
1989	35 (35)	28 (27)	19 (20)	15 (14)	31 (32)	25 (24)
1990	37 (36)	29 (28)	20 (19)	15 (14)	33 (32)	26 (25)
1991	35 (35)	28 (27)	19 (19)	14 (14)	31 (31)	25 (24)
1992	33 (32)	26 (25)	18 (18)	13 (13)	30 (29)	23 (22)
1993	32 (31)	25 (24)	17 (17)	13 (12)	28 (27)	22 (21)
1994	30 (29)	24 (22)	17 (15)	13 (11)	27 (25)	21 (19)
1995	30 (28)	23 (22)	16 (16)	12 (11)	26 (25)	20 (19)
1996	29 (28)	23 (21)	16 (16)	12 (11)	26 (25)	20 (19)
1997	28 (27)	22 (20)	15 (14)	11 (10)	24 (24)	19 (18)
1998	28 (27)	22 (20)	15 (14)	11 (10)	24 (23)	19 (18)
1999	28 (26)	21 (20)	14 (14)	11 (10)	24 (23)	19 (18)
2000	29 (27)	22 (20)	16 (16)	12 (11)	26 (24)	20 (18)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
(values in parentheses represent estimates from the ***old*** *imputation methodology)*

Table 4 shows the alcohol involvement of drivers involved in fatal crashes by the type of vehicle they were driving. Overall, the highest proportion of drivers involved who had any alcohol were motorcycle operators. The lowest such proportion is observed with drivers of large trucks.

Table 4 : Alcohol Involvement for Drivers Involved in Fatal Crashes by Vehicle Type and Driver's BAC, FARS 1982-2000*

Year	Passenger Cars		Light Trucks and Vans		Large Trucks		Motorcycles	
	.01+	.10+	.01+	.10+	.01+	.10+	.01+	.10+
1982	42 (40)	33 (31)	44 (43)	36 (35)	10 (8)	5 (4)	55 (53)	43 (41)
1983	40 (39)	33 (30)	43 (42)	36 (33)	10 (8)	6 (5)	57 (54)	43 (41)
1984	39 (36)	30 (28)	41 (39)	32 (31)	9 (8)	6 (4)	55 (54)	42 (40)
1985	36 (35)	28 (26)	37 (36)	30 (29)	7 (6)	4 (4)	53 (53)	39 (39)
1986	36 (35)	28 (26)	38 (37)	30 (29)	7 (5)	4 (3)	56 (54)	42 (41)
1987	35 (34)	27 (25)	37 (37)	29 (29)	5 (4)	3 (3)	51 (51)	39 (38)
1988	34 (33)	26 (25)	37 (37)	29 (29)	6 (5)	3 (3)	51 (50)	37 (36)
1989	32 (32)	25 (24)	35 (35)	28 (28)	4 (5)	3 (3)	53 (53)	41 (40)
1990	34 (32)	27 (24)	36 (36)	29 (29)	5 (5)	2 (2)	52 (52)	40 (39)
1991	31 (31)	25 (23)	35 (36)	28 (28)	4 (4)	2 (2)	52 (51)	40 (39)
1992	30 (29)	23 (22)	33 (33)	26 (26)	3 (3)	2 (1)	49 (48)	37 (36)
1993	28 (27)	22 (21)	31 (31)	25 (25)	4 (3)	2 (2)	45 (44)	34 (33)
1994	28 (26)	22 (19)	29 (29)	23 (23)	3 (3)	2 (1)	41 (40)	30 (29)
1995	27 (26)	21 (19)	29 (28)	23 (22)	4 (3)	2 (1)	42 (41)	30 (29)
1996	27 (26)	21 (19)	28 (28)	22 (22)	3 (3)	2 (1)	43 (42)	32 (30)
1997	26 (24)	20 (18)	26 (26)	21 (20)	3 (2)	1 (1)	41 (39)	30 (28)
1998	26 (24)	20 (18)	26 (26)	21 (20)	2 (2)	1 (1)	41 (40)	32 (30)
1999	25 (24)	19 (18)	26 (26)	21 (20)	3 (2)	1 (1)	40 (38)	29 (28)
2000	28 (26)	22 (19)	26 (26)	20 (20)	3 (2)	1 (1)	40 (38)	29 (27)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
*(values in parentheses represent estimates from the **old** imputation methodology)*

Table 5 shows the extent of alcohol involvement of drivers involved in fatal crashes by their age.

Table 5 : Alcohol Involvement for Drivers Involved in Fatal Crashes by Age and Driver's BAC, FARS 1982-2000*

Age	**16-20**		**21-24**		**25-34**		**35-44**		**45-64**		**65+**	
BAC	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**
1982	45 (44)	32 (31)	53 (52)	42 (40)	46 (44)	38 (35)	38 (35)	31 (28)	29 (26)	23 (21)	15 (14)	11 (10)
1983	43 (42)	31 (30)	53 (51)	42 (39)	46 (44)	38 (35)	37 (34)	31 (28)	26 (25)	22 (19)	13 (12)	9 (9)
1984	40 (40)	28 (27)	52 (49)	40 (37)	44 (42)	36 (33)	35 (32)	29 (26)	25 (23)	20 (18)	14 (12)	10 (9)
1985	35 (35)	23 (24)	47 (46)	37 (35)	42 (41)	34 (32)	32 (30)	27 (24)	23 (22)	18 (17)	12 (11)	8 (8)
1986	37 (36)	25 (24)	49 (47)	38 (36)	43 (41)	35 (33)	33 (31)	27 (25)	23 (21)	18 (16)	12 (10)	7 (7)
1987	33 (33)	22 (21)	47 (45)	36 (34)	43 (42)	34 (33)	32 (31)	27 (25)	21 (21)	17 (16)	11 (10)	7 (7)
1988	33 (32)	22 (21)	47 (46)	36 (35)	42 (41)	34 (33)	32 (31)	27 (25)	21 (21)	17 (16)	11 (11)	7 (7)
1989	30 (30)	20 (20)	45 (45)	35 (35)	40 (40)	33 (32)	32 (31)	27 (25)	21 (21)	17 (17)	10 (10)	6 (7)
1990	33 (32)	22 (21)	46 (45)	36 (35)	43 (41)	35 (33)	33 (32)	28 (26)	21 (20)	17 (16)	10 (10)	7 (6)
1991	30 (30)	21 (20)	45 (44)	35 (34)	41 (40)	34 (32)	32 (31)	27 (25)	20 (20)	16 (16)	9 (10)	6 (6)
1992	27 (27)	18 (18)	42 (41)	32 (31)	40 (38)	32 (31)	31 (30)	26 (24)	20 (19)	15 (14)	10 (9)	6 (6)
1993	24 (25)	16 (16)	40 (39)	31 (31)	37 (36)	30 (29)	30 (29)	25 (23)	20 (19)	16 (14)	8 (8)	6 (5)
1994	24 (23)	15 (14)	39 (37)	30 (28)	36 (34)	29 (27)	29 (27)	24 (22)	19 (17)	15 (14)	9 (8)	6 (5)
1995	21 (21)	13 (13)	38 (37)	29 (28)	35 (34)	28 (27)	30 (29)	24 (23)	19 (18)	15 (14)	8 (7)	5 (5)
1996	23 (21)	15 (14)	38 (37)	28 (27)	34 (33)	28 (26)	29 (28)	24 (22)	19 (18)	15 (14)	9 (8)	6 (5)
1997	22 (22)	15 (14)	36 (35)	27 (26)	32 (31)	25 (24)	29 (27)	24 (22)	18 (17)	14 (13)	8 (7)	5 (5)
1998	22 (22)	15 (14)	37 (36)	29 (28)	32 (31)	25 (24)	28 (27)	23 (21)	18 (17)	14 (13)	8 (7)	5 (5)
1999	22 (21)	15 (14)	38 (36)	28 (27)	32 (30)	25 (24)	28 (27)	23 (21)	18 (17)	14 (13)	8 (7)	5 (4)
2000	24 (23)	16 (15)	38 (37)	29 (27)	33 (31)	26 (24)	30 (28)	24 (22)	19 (18)	15 (14)	8 (8)	5 (5)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
*(values in parentheses represent estimates from the **old** imputation methodology)*

Drivers in the age group 21-24 have the highest proportion of alcohol involvement. This trend was carried through 1982 to 1999. Older drivers, 65 years of age and above had the lowest proportion of alcohol involvement.

Table 6 shows the alcohol involvement among drivers who are fatally injured. Classification of this population by time of day shows that a greater proportion of drivers killed in nighttime crashes had alcohol involvement than drivers that are killed in daytime crashes.

Table 6 : Alcohol Involvement for Drivers Killed in Fatal Crashes by Time of the Day and Driver's BAC, FARS 1982-2000*

Year	Daytime		Nighttime		Total	
	.01+	.10+	.01+	.10+	.01+	.10+
1982	28 (26)	22 (20)	73 (70)	62 (59)	55 (53)	46 (44)
1983	26 (25)	20 (19)	73 (70)	62 (59)	54 (51)	45 (42)
1984	25 (23)	19 (17)	71 (69)	59 (57)	51 (49)	42 (40)
1985	24 (23)	17 (17)	69 (68)	57 (56)	49 (48)	39 (39)
1986	24 (23)	18 (16)	70 (68)	58 (56)	50 (48)	40 (39)
1987	23 (22)	17 (16)	67 (66)	56 (55)	48 (47)	39 (38)
1988	22 (22)	16 (16)	67 (67)	56 (56)	47 (47)	38 (38)
1989	21 (21)	16 (15)	67 (66)	56 (56)	46 (46)	38 (37)
1990	21 (21)	16 (15)	67 (67)	57 (57)	46 (46)	38 (38)
1991	20 (19)	15 (14)	66 (65)	56 (55)	45 (44)	37 (37)
1992	19 (18)	14 (13)	64 (63)	54 (53)	43 (42)	35 (34)
1993	18 (17)	13 (12)	63 (62)	54 (52)	41 (40)	34 (33)
1994	17 (16)	13 (12)	60 (60)	51 (50)	38 (37)	31 (31)
1995	18 (17)	13 (12)	61 (59)	51 (50)	39 (38)	32 (31)
1996	17 (16)	12 (11)	61 (59)	51 (50)	38 (37)	31 (30)
1997	16 (15)	12 (11)	58 (57)	49 (48)	36 (35)	30 (28)
1998	16 (15)	12 (11)	59 (57)	49 (48)	36 (35)	30 (28)
1999	17 (15)	12 (11)	58 (56)	49 (47)	36 (35)	29 (28)
2000	17 (16)	12 (11)	58 (57)	49 (48)	37 (36)	30 (29)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
*(values in parentheses represent estimates from the **old** imputation methodology)*

Table 7 shows the trend of alcohol involvement among fatally injured drivers by the time of day and the type of crash, i.e., if the driver was killed in a single vehicle or a multiple vehicle crash.

Table 7 : Alcohol Involvement for Drivers Killed in Fatal Crashes by Crash Type and Time of the Day and Driver's BAC, FARS 1982-2000*

	Single Vehicle Crashes						Multiple Vehicle Crashes					
Time	**Day**		**Night**		**Total**		**Day**		**Night**		**Total**	
BAC	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**
1982	42 (40)	34 (32)	83 (81)	73 (71)	71 (69)	61 (60)	20 (18)	14 (12)	59 (56)	47 (43)	40 (37)	31 (28)
1983	40 (38)	33 (31)	83 (80)	73 (71)	69 (67)	60 (58)	18 (17)	13 (11)	59 (56)	47 (43)	38 (36)	29 (27)
1984	37 (36)	30 (28)	81 (79)	70 (69)	67 (65)	58 (56)	18 (16)	12 (11)	56 (54)	43 (41)	35 (34)	26 (24)
1985	37 (36)	29 (28)	80 (79)	68 (68)	65 (64)	55 (55)	17 (16)	11 (10)	54 (53)	42 (40)	33 (32)	25 (24)
1986	36 (34)	29 (27)	80 (79)	69 (67)	66 (64)	56 (54)	18 (17)	12 (11)	54 (52)	42 (40)	34 (33)	25 (24)
1987	36 (35)	29 (28)	78 (77)	67 (66)	64 (63)	55 (54)	16 (15)	10 (10)	52 (51)	40 (39)	32 (31)	24 (23)
1988	34 (34)	27 (27)	78 (78)	68 (68)	63 (63)	54 (54)	16 (15)	10 (9)	51 (51)	40 (39)	31 (31)	23 (22)
1989	34 (34)	27 (27)	77 (76)	67 (66)	62 (62)	53 (53)	15 (14)	10 (9)	52 (51)	41 (40)	30 (30)	23 (22)
1990	33 (33)	26 (27)	78 (78)	68 (68)	63 (63)	54 (54)	14 (14)	9 (9)	51 (50)	40 (40)	30 (29)	23 (22)
1991	31 (31)	25 (25)	77 (76)	67 (67)	62 (61)	53 (52)	13 (13)	9 (8)	49 (48)	39 (38)	28 (27)	21 (20)
1992	30 (29)	23 (23)	76 (74)	66 (65)	59 (58)	51 (50)	13 (12)	8 (7)	46 (45)	37 (35)	26 (25)	20 (19)
1993	29 (28)	23 (23)	74 (73)	65 (64)	58 (57)	50 (49)	12 (11)	8 (7)	46 (45)	37 (35)	25 (24)	19 (18)
1994	27 (26)	22 (21)	72 (71)	63 (62)	55 (54)	47 (46)	11 (11)	7 (7)	44 (43)	35 (34)	24 (23)	18 (17)
1995	28 (27)	22 (22)	73 (71)	63 (62)	56 (55)	48 (47)	12 (11)	8 (7)	42 (41)	34 (32)	23 (22)	18 (16)
1996	27 (26)	21 (21)	73 (71)	63 (62)	55 (54)	47 (46)	11 (11)	7 (6)	44 (42)	34 (32)	23 (22)	17 (16)
1997	26 (24)	21 (19)	70 (69)	61 (60)	53 (51)	45 (44)	11 (10)	7 (6)	42 (39)	33 (31)	22 (21)	17 (15)
1998	26 (25)	21 (20)	70 (68)	61 (60)	53 (51)	45 (44)	11 (10)	7 (6)	40 (38)	31 (29)	21 (20)	16 (14)
1999	26 (25)	20 (19)	70 (68)	60 (59)	52 (51)	44 (43)	11 (10)	7 (6)	40 (38)	31 (28)	21 (20)	15 (14)
2000	26 (25)	20 (20)	71 (69)	60 (59)	53 (51)	44 (43)	11 (10)	7 (6)	40 (39)	32 (30)	22 (21)	16 (15)

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*
(values in parentheses represent estimates from the ***old*** *imputation methodology)*

The trend of alcohol involvement is similar for both methods with the nighttime single vehicle crashes showing the highest proportion of drivers that had some alcohol.

Table 8: Alcohol Involvement for Drivers Killed in Fatal Crashes by Time of the Day and Day of the Week and Driver's BAC, FARS 1982-2000*

	Weekday						Weekend					
Time	**Day**		**Night**		**Total**		**Day**		**Night**		**Total**	
BAC	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**	**.01+**	**.10+**
1982	24 (22)	19 (16)	69 (66)	58 (56)	46 (43)	38 (35)	38 (37)	30 (28)	76 (74)	65 (62)	67 (65)	56 (54)
1983	22 (21)	17 (15)	69 (67)	59 (56)	44 (42)	36 (34)	38 (35)	30 (27)	76 (73)	65 (62)	66 (63)	56 (53)
1984	21 (19)	15 (14)	66 (63)	55 (53)	41 (39)	33 (31)	35 (34)	27 (26)	75 (73)	62 (61)	64 (62)	53 (51)
1985	20 (19)	14 (13)	64 (63)	53 (53)	39 (38)	31 (30)	36 (35)	27 (26)	73 (72)	60 (60)	62 (61)	51 (50)
1986	21 (19)	15 (14)	65 (63)	54 (53)	41 (39)	32 (31)	34 (32)	25 (23)	73 (71)	60 (59)	62 (60)	50 (49)
1987	19 (18)	14 (13)	63 (62)	52 (51)	38 (37)	30 (29)	34 (33)	26 (25)	71 (70)	59 (58)	60 (59)	49 (48)
1988	18 (18)	13 (13)	62 (62)	52 (52)	37 (37)	30 (29)	33 (32)	25 (24)	71 (71)	60 (59)	60 (60)	50 (49)
1989	17 (17)	12 (12)	62 (61)	52 (52)	35 (35)	29 (29)	33 (32)	25 (24)	71 (70)	59 (59)	60 (59)	50 (49)
1990	17 (17)	12 (12)	62 (62)	53 (52)	36 (35)	29 (29)	31 (31)	24 (24)	71 (70)	60 (60)	60 (59)	50 (50)
1991	16 (16)	12 (11)	62 (61)	53 (52)	35 (34)	29 (28)	30 (29)	24 (23)	70 (69)	59 (58)	58 (57)	49 (48)
1992	16 (15)	11 (10)	58 (58)	49 (49)	33 (32)	26 (26)	29 (27)	22 (20)	69 (67)	58 (57)	57 (55)	47 (46)
1993	15 (14)	11 (10)	57 (56)	48 (48)	31 (30)	25 (25)	27 (26)	20 (19)	68 (66)	58 (56)	55 (53)	46 (44)
1994	14 (13)	10 (9)	53 (52)	44 (43)	28 (27)	22 (22)	25 (24)	20 (19)	67 (65)	57 (56)	53 (52)	45 (43)
1995	14 (14)	10 (10)	54 (53)	46 (45)	29 (29)	24 (23)	27 (25)	20 (19)	66 (64)	56 (54)	53 (51)	44 (42)
1996	14 (13)	10 (9)	54 (53)	45 (44)	29 (28)	23 (22)	26 (24)	20 (18)	67 (65)	56 (55)	53 (51)	44 (42)
1997	13 (12)	9 (8)	51 (50)	43 (42)	27 (26)	22 (21)	25 (23)	19 (17)	64 (62)	54 (53)	51 (48)	42 (40)
1998	13 (12)	10 (9)	52 (50)	43 (42)	27 (26)	22 (21)	24 (23)	19 (17)	65 (62)	55 (53)	50 (48)	42 (40)
1999	13 (12)	9 (8)	51 (50)	43 (42)	27 (26)	21 (20)	27 (25)	20 (18)	64 (61)	53 (51)	50 (48)	41 (39)
2000	13 (13)	9 (9)	51 (50)	42 (41)	27 (26)	21 (21)	25 (23)	19 (18)	65 (63)	54 (53)	51 (49)	42 (40)

Table 9 shows the level of alcohol involvement among fatally injured drivers by their restraint use. Restraint usage has been observed to be correlated with the level of alcohol involvement and is also used as a covariate in both the models.

Table 9: Alcohol Involvement of Unbelted, Fatally Injured Drivers of Passenger Cars, Light Trucks and Vans and Driver's BAC, FARS 1982-2000*

Year	BAC=.01+		BAC=.10+	
	Multiple Imputation	Discriminant Analysis	Multiple Imputation	Discriminant Analysis
1982	58	55	49	47
1983	55	53	48	45
1984	53	51	45	43
1985	52	51	43	42
1986	55	53	46	44
1987	54	53	45	44
1988	55	55	46	46
1989	53	53	45	45
1990	55	54	46	46
1991	54	53	46	45
1992	52	51	44	43
1993	51	49	44	42
1994	49	48	42	41
1995	50	48	42	41
1996	50	48	42	40
1997	48	46	41	39
1998	48	46	41	39
1999	47	45	40	38
2000	49	47	41	39

*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*

Table 10 shows the level of alcohol involvement among fatally injured drivers who had prior convictions such as previously recorded crashes, DWI convictions, Speeding and License suspensions. Drivers who had previous DWI convictions who were killed in crashes were more likely to have alcohol compared to drivers with a history of other types of infractions.

Table 10: Previous Driving Records of Drivers Killed in Traffic Crashes, FARS 1982-2000*

Year	Recorded Crashes		DWI Convictions		Speeding Convictions		Recorded Suspensions	
	.01+	.10+	.01+	.10+	.01+	.10+	.01+	.10+
1982	62 (59)	52 (50)	85 (83)	76 (74)	64 (61)	53 (50)	77 (75)	67 (65)
1983	60 (58)	51 (48)	84 (82)	74 (73)	62 (60)	52 (49)	75 (73)	65 (63)
1984	59 (58)	48 (47)	83 (82)	73 (72)	59 (57)	48 (46)	72 (69)	61 (59)
1985	54 (53)	44 (44)	82 (81)	73 (72)	56 (56)	46 (45)	73 (71)	63 (62)
1986	55 (54)	44 (43)	83 (82)	74 (73)	57 (56)	46 (45)	75 (73)	64 (62)
1987	52 (52)	42 (41)	82 (81)	74 (73)	54 (54)	44 (43)	71 (70)	61 (60)
1988	52 (52)	43 (43)	83 (83)	75 (75)	55 (55)	45 (45)	70 (70)	60 (60)
1989	51 (51)	42 (42)	82 (82)	74 (73)	54 (53)	43 (43)	69 (69)	60 (60)
1990	50 (50)	41 (41)	83 (83)	76 (76)	53 (53)	44 (44)	70 (70)	60 (60)
1991	50 (49)	42 (41)	84 (83)	77 (76)	54 (53)	44 (44)	70 (69)	61 (60)
1992	46 (46)	38 (38)	84 (84)	76 (76)	49 (48)	40 (39)	68 (67)	59 (58)
1993	45 (44)	37 (36)	80 (79)	73 (72)	47 (46)	39 (38)	67 (66)	58 (58)
1994	41 (40)	33 (32)	81 (81)	74 (73)	44 (43)	35 (34)	64 (63)	55 (54)
1995	41 (41)	34 (33)	81 (80)	74 (74)	46 (45)	37 (36)	64 (63)	55 (54)
1996	40 (40)	32 (32)	79 (79)	72 (71)	45 (44)	36 (35)	63 (62)	54 (53)
1997	40 (39)	32 (31)	78 (77)	70 (69)	43 (42)	35 (34)	63 (62)	55 (53)
1998	40 (38)	32 (31)	76 (75)	69 (68)	43 (42)	34 (33)	62 (60)	53 (52)
1999	37 (36)	31 (30)	78 (78)	70 (69)	43 (41)	34 (33)	61 (60)	52 (51)
2000	40 (38)	32 (31)	79 (77)	70 (70)	42 (41)	34 (33)	62 (60)	52 (50)

*(values in parentheses represent estimates from the **old** imputation methodology)*
*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*

Table 11 shows the level of alcohol involvement among fatally injured nonoccupants (pedestrians and pedalcyclists).

Table 11 : Pedestrians and Pedalcyclists Killed in Traffic Crashes by BAC of the Pedestrian or Pedalcyclist, FARS 1982-2000*

Year	Pedestrians		Pedalcyclists	
	.01+	.10+	.01+	.10+
1982	42 (41)	36 (34)	22 (20)	16 (14)
1983	42 (40)	36 (33)	18 (20)	14 (14)
1984	40 (39)	34 (33)	18 (18)	13 (13)
1985	40 (39)	33 (32)	15 (18)	11 (12)
1986	39 (39)	33 (32)	17 (18)	13 (12)
1987	38 (38)	31 (31)	19 (21)	14 (14)
1988	37 (37)	31 (30)	18 (19)	14 (14)
1989	39 (39)	32 (32)	18 (19)	14 (14)
1990	38 (38)	32 (32)	20 (21)	16 (16)
1991	38 (38)	32 (32)	24 (24)	18 (17)
1992	39 (38)	33 (32)	20 (22)	15 (16)
1993	38 (37)	32 (32)	22 (23)	17 (17)
1994	36 (36)	30 (30)	20 (21)	16 (16)
1995	37 (37)	31 (30)	23 (24)	19 (19)
1996	38 (38)	32 (32)	22 (23)	17 (17)
1997	35 (34)	30 (29)	22 (23)	17 (17)
1998	38 (37)	31 (30)	24 (24)	19 (19)
1999	38 (37)	32 (31)	26 (26)	23 (22)
2000	38 (37)	32 (30)	25 (26)	21 (21)

*(values in parentheses represent estimates from the **old** imputation methodology)*
*(*Based on 1982-1999 Final and 2000 Annual Report (AR) Files)*

Table 12 compares the estimates from the two methods for all traffic fatalities by state and the highest BAC in the crash for 2000.

Table 12 : Percentage Alcohol Involvement in Traffic Fatalities by State and Highest BAC in the Crash, FARS 2000*

State	BAC=.01+	BAC=.10+	State	BAC=.01+	BAC=.10+
Alabama	43 (40)	35 (33)	Montana	49 (46)	42 (39)
Alaska	54 (52)	45 (43)	Nebraska	38 (37)	26 (25)
Arizona	45 (44)	36 (34)	Nevada	43 (45)	35 (35)
Arkansas	34 (31)	25 (21)	New Hampshire	37 (39)	30 (31)
California	39 (37)	30 (28)	New Jersey	44 (44)	35 (32)
Colorado	40 (38)	31 (29)	New Mexico	49 (48)	39 (37)
Connecticut	47 (46)	38 (35)	New York	32 (29)	23 (20)
Delaware	50 (49)	41 (40)	North Carolina	39 (36)	32 (28)
DC	38 (39)	31 (29)	North Dakota	49 (48)	43 (42)
Florida	43 (40)	34 (31)	Ohio	41 (38)	34 (30)
Georgia	38 (37)	30 (28)	Oklahoma	36 (34)	28 (26)
Hawaii	43 (41)	31 (28)	Oregon	41 (42)	29 (29)
Idaho	43 (41)	30 (29)	Pennsylvania	43 (41)	36 (34)
Illinois	44 (43)	35 (34)	Rhode Island	52 (51)	40 (38)
Indiana	34 (31)	27 (24)	South Carolina	41 (40)	34 (31)
Iowa	30 (28)	25 (22)	South Dakota	48 (47)	39 (38)
Kansas	35 (33)	27 (26)	Tennessee	43 (39)	34 (31)
Kentucky	34 (31)	27 (25)	Texas	49 (50)	40 (38)
Louisiana	49 (48)	39 (38)	Utah	28 (24)	22 (18)
Maine	31 (30)	23 (22)	Vermont	41 (39)	36 (34)
Maryland	40 (38)	30 (27)	Virginia	38 (37)	29 (28)
Massachusetts	49 (50)	38 (35)	Washington	45 (44)	36 (34)
Michigan	38 (37)	30 (29)	West Virginia	45 (43)	38 (36)
Minnesota	41 (41)	33 (33)	Wisconsin	44 (43)	37 (36)
Mississippi	41 (40)	32 (30)	Wyoming	32 (30)	28 (26)
Missouri	44 (44)	36 (33)	**U.S. Total**	**41 (40)**	**33 (31)**

*(values in parentheses represent estimates from the **old** imputation methodology)*
*(*Based on 2000 Annual Report (AR) File)*

6. Conclusions

NHTSA is adopting the Multiple Imputation procedure for estimating missing BAC values for the significant analytical advantages it provides over the Discriminant Analysis. There is a discrepancy in the estimates between the two methods with the estimates from Multiple Imputation providing estimates that are up to 2 percent higher than those provided by Discriminant Analysis. The overall trend of alcohol involvement tracks very similarly for the estimates from both methods. The historical revision of estimates using Multiple Imputation up to 1982 will preserve the overall trend of alcohol involvement. The discrepancy between the two methods can be attributed to the fact that Multiple Imputation uses the logit model as compared to the linear discriminant model of the old method. Fundamental differences in the assumptions involved in the two methods can be one of the main reasons for the shift in estimates.

This underscores the importance of providing meaningful estimates of uncertainty (e.g. standard errors) for statistics related to BAC. The standard errors now available from Multiple Imputation will enable NCSA to provide measures of uncertainty. Also, the BAC values arrived at through Multiple Imputation can now be used as a factor in analytical models.

7. Transitioning Schedule

NCSA will use the old method to report alcohol involvement for the 2001 Early Assessment.

NCSA will use Multiple Imputation to report alcohol involvement for the 2001 FARS Annual Report.

> The new estimates will be used in NCSA's Annual Publications (Traffic Safety Facts, Fact Sheets etc.), as well as related Reports and Research Notes that use the 2001 FARS data.
>
> All historical series of alcohol involvement in these publications will be revised back to the 1982 data year to reflect the estimates from the new methodology.

The revised alcohol estimates for prior years will differ from the estimates that are in previously published reports for those years, the extent of which has been documented in this report.

NCSA will also make the new datasets available on its website to enable non-NCSA users to generate estimated alcohol involvement along various categories of interest.

A web-interface to the new estimates is being implemented and should be online by the time the public-use datasets are released.

8. References

1. De Wolf, V.A. (1988) *Panel Meeting on the Imputation of Missing Data in FARS: Summary and Recommendations*, National Highway Traffic Safety Administration, Department of Transportation.

2. Huberty, C.J. and Wisenbaker, J.M. (1987) *On Estimating Missing BAC Measures in the 1985 FARS Data Set*, National Highway Traffic Safety Administration, Department of Transportation.

3. Kalton G. (1988) *Comments on Imputation for Missing Data in the Fatal Accident Reporting System*, National Highway Traffic Safety Administration, Department of Transportation.

4. Klein, T.M., (1986) *A Method for estimating Posterior BAC distributions for persons involved in fatal traffic accidents.* Report DOT-HS-807-094, National Highway Traffic Safety Administration, Department of Transportation.

5. Little, R. (1987) *Imputation Methodology for the Fatal Accident Reporting System*, National Highway Traffic Safety Administration, Department of Transportation.

6. Rubin, D.B. (1987) *Multiple Imputation of Nonresponse in Surveys.* J.Wiley and Sons, New York.

7. Rubin, D.B. (1987) *Imputation of Missing Data in FARS: Suggestions and Comments on Papers and Presentations,* National Highway Traffic Safety Administration, Department of Transportation.

8. Rubin,D.B, Shafer, J.L. and Subramanian, R. (1998) *Multiple Imputation of Missing Blood Alcohol Concentration (BAC) values in FARS.* Report DOT-HS-808-816, National Highway Traffic Safety Administration, Department of Transportation.

Appendix A: FAQ on the Multiply-Imputed Datasets of Missing BAC in FARS

1. What is imputation?

   A. Imputation is the practice of 'filling in' missing data with plausible values. It solves the missing-data problem at the beginning of the analysis.

2. Why impute Missing BAC in FARS?

   A. On an average, approximately 60 percent of the BAC values are missing/unknown in FARS each year. Invalid inferences can be drawn on the level of alcohol involvement for cases where the BAC is missing as the characteristics of the persons with unknown BACs can be significantly different from those with known BACs. In order to perform complete-data analysis of FARS data with respect to alcohol involvement, the missing BACs need to be simulated (imputation!)

3. What is Multiple Imputation (MI)?

   A. MI is a technique in which each missing value is replaced by m>1 simulated versions and these simulated complete datasets are analyzed by standard methods. These simulated values are actual values of BAC in the plausible range (.00<=BAC<=.94).

4. Why Multiple Imputation of BAC in FARS?

   A. Multiple Imputation is the state-of-the-art technique to impute missing values. Each missing BAC value is replaced by ten simulated values of BAC using rigorous statistical techniques that consider the interaction of all the characteristics of the case. MI allows for the computation of Standard Errors and Confidence Intervals.

5. Can MI estimates be used in analysis (regression etc.)?

   A. Yes, the multiply-imputed values can be used in analysis. The regression coefficients will have to be averaged out over the ten imputed values of BAC.

6. How do I combine the results across the multiply imputed datasets?

   A. The data analysis for the quantity of interest (e.g., Percent Alcohol Involvement for Drivers involved in Fatal Crashes), or D_{inv} , should be performed ten times, once for each of the ten imputed datasets, to obtain a single set of results.
   From each analysis, suppose that D^{j}_{inv} is the percent of alcohol involvement for drivers involved from the j^{th} imputed dataset. The overall estimate for drivers involved will be average of the individual estimate from the ten datasets.

$$D_{inv} = \frac{1}{m}\sum_{j=1}^{10} D_{inv}^{j}$$

7. Why not just impute once?

   A. If the proportion of missing values is small, then single imputation may be quite reasonable. Without special corrective measures, single-imputation inference tends to overstate precision because it omits the between-imputation component of variability which is the error in estimating the missing value. When the fraction of missing information is small (say, less than 5%) then single-imputation inferences may be fairly accurate, which is not the case with FARS, where more than 50 percent of the BAC values are missing. For joint inferences about multiple variables, however, even small rates of missing information may seriously impair a single-imputation procedure.

8. Will the alcohol involvement estimates change from those of the previous method?

   A. Yes, there will be minor differences between the estimates of alcohol involvement between the earlier method (Discriminant Analysis) and Multiple Imputation. The MI estimates are overall between 0 to 2 percent higher than the estimates from the old methodology.

9. Why are there differences between the results from the two methods?

   A. The imputation methodologies have different statistical models to estimate missing BAC values that could lead to the observed differences in the estimates. The old method computes probabilities of involvement along definite categories of BAC while MI imputes actual values of BAC.

10. Are there sample programs that analyze the multiply imputed datasets?

    A. Yes, there are sample programs written in the SAS programming language that compute point-estimates and the standard errors which are documented in the following section. Also, SAS® has released a trial version of PROC MIANALYZE® to analyze multiply-imputed datasets. This procedure should have packaged routines to generate descriptive statistics and point-estimates from the multiply-imputed datasets.

Appendix B: Sample SAS programs to analyze the Multiply-imputed Datasets in FARS

Example 1: Program to determine the extent of alcohol involvement in fatalities

(1) This section of code creates a dataset **CRASHES** which is a result of a merge between the crash level file for 1999 and the multiple imputation dataset (**MIACC99**). The dataset retains the ten imputations (**A1** to **A10**) and **FATALS** which will be used later in the tabulation procedures.

```
DATA CRASHES;
   MERGE FARS99.ACCIDENT (IN=A KEEP=ST_CASE FATALS) FARS99.MIACC99 (IN=B);
IF A AND B;
BY ST_CASE;
OUTPUT;
RUN;
```

(2) The first step in the next segment is the creation of **CRASHBAC** that has for every record **fpc1..fpc10**, **spc1..spc10** and **tpc1..tpc10. fpc** is the indicator variable for the first category namely, BAC=0. Hence if the first imputation is equal to 0 then **fpc1** is set to 1, or, if the first imputation is 8 then **spc1** is set to 1, or, if the first imputation is equal to 15 then **tpc1** is set to 1. The macro ***mi*** runs through each of the ten imputations and sets the values of **fpc&i**, **spc&i** and **tpc&i** to 0 or 1 depending upon the value of the imputation. Note that the imputed values are scaled values of BAC by a factor of 100, i.e., 10 actually corresponds to a BAC value of .10.

```
data CRASHBAC;
set CRASHES;
%macro mi;
%do i=1 %to 10;
if A&i=0 then fpc&i=1;
else fpc&i=0;
if (1<=A&i<=9) then spc&i=1;  /* Use (1<=A&i<=7) for .08 Analyses */
else spc&i=0;
if (A&i>=10) then tpc&i=1;    /* Use (A&i>=8) for .08 Analyses          */
else tpc&i=0;
%end;
%mend mi;
%mi;
RUN;
```

(4) The next segment is the procedure **means** that computes the sum of **fpc&i,spc&i and tpc&i** and stores them in a dataset **case&i** for every imputation. Thus ten temporary datasets are created containing the sums **fsbac&i, ssbac&i and tsbac&i**.

```
%macro DO_MEANS;
%do i=1 %to 10;
proc means noprint data=crashbac;
var fpc&i spc&i tpc&i;
freq FATALS;            /* WEIGH EACH CRASH BY NUMBER OF FATALITIES    */
output out=case&i n=total sum=fsbac&i ssbac&i tsbac&i;
run;
%end;
%mend DO_MEANS;
%DO_MEANS;
run;
```

(5) The next section of the code combines the ten datasets **case1**..**case10** and computes the Multiple-imputation point estimate $\overline{P}$ for each interval of study. For example, **pcnt0** is the $\overline{P}$ for the first interval of study, namely BAC=0.

```
data mi_est;
%macro AVG_EST;
%do i=1 %to 10;
set case&i;
sbac0=mean(fsbac1,fsbac2,fsbac3,fsbac4,fsbac5,fsbac6,fsbac7,fsbac8,fsbac9,fsb
ac10);
sbac1=mean(ssbac1,ssbac2,ssbac3,ssbac4,ssbac5,ssbac6,ssbac7,ssbac8,ssbac9,ssb
ac10);
sbac2=mean(tsbac1,tsbac2,tsbac3,tsbac4,tsbac5,tsbac6,tsbac7,tsbac8,tsbac9,tsb
ac10);
sbac3=sbac1+sbac2;
%end;
%mend AVG_EST;
%AVG_EST;
run;
```

(6) The next section of code tabulates the alcohol involvement percentages across the three categories and prints them to the output.

```
PROC TABULATE DATA=MI_EST FORMAT=COMMA10.0 MISSING;
VAR SBAC0 SBAC1 SBAC2 SBAC3 TOTAL;
TABLE (SBAC0='BAC=.00' SBAC1='BAC=.01-.09' SBAC2='BAC=.10+'
            TOTAL='Total Fatalities'
            SBAC3='Alcohol-Related Fatalities (0.01+)')*
            (SUM PCTSUM<TOTAL>) / RTS=15;
KEYLABEL N=' ' ALL='Total' SUM='Number' PCTSUM='Percent';
TITLE1 'FATALITIES BY EXTENT OF ALCOHOL INVOLVEMENT';
TITLE2 'FARS  1999';
RUN;
```

The TABULATE procedure outputs a Table as shown below.

FATALITIES BY EXTENT OF ALCOHOL INVOLVEMENT
FARS 1999

BAC=.00		BAC=.01-.09		BAC=.10+		Total Fatalities		Alcohol-Related Fatalities (0.01+)	
Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
25,145	60	3,391	8	13,181	32	41,717	100	16,572	40

Example 2: Program to determine alcohol involvement for drivers involved in fatal crashes in FARS by the sex of the driver

(1) This section of code creates a dataset **DRVINV** which is a result of a merge between the person level file for 1999 and the multiple imputation dataset (**MIPER99**). The dataset retains the ten imputations (**P1** to **P10**) and **SEX** which will be used later in the tabulation procedures. **SEX** is 1 in the case of a male driver and 2 for a female driver and 9 when it is not known.

```
DATA DRVINV;
      MERGE FARS99.PERSON (IN=A KEEP=ST_CASE VEH_NO PER_NO SEX PER_TYP)
      FARS99.MIPER99 (IN=B);
IF A AND B;
IF PER_TYP=1;
BY ST_CASE VEH_NO PER_NO;
OUTPUT;
RUN;
```

(2) The first step in the next segment is the creation of **DRVBAC** that has for every record **fpc1..fpc10**, **spc1..spc10** and **tpc1..tpc10. fpc** is the indicator variable for the first category namely, BAC=0. Hence if the first imputation is equal to 0 then **fpc1** is set to 1, or, if the the first imputation is 8 then **spc1** is set to 1, or, if the first imputation is equal to 15 then **tpc1** is set to 1. The macro ***mi*** runs through each of the ten imputations and sets the values of **fpc&i**, **spc&i** and **tpc&i** to 0 or 1 depending upon the value of the imputation. Note that the imputed values are scaled values of BAC by a factor of 100, i.e., 10 actually corresponds to a BAC value of .10.

```
data DRVBAC;
set DRVINV;
%macro mi;
%do i=1 %to 10;
if P&i=0 then fpc&i=1;
else fpc&i=0;
if (1<=P&i<=9) then spc&i=1;  /* Use (1<=P&i<=7) for .08 Analyses */
else spc&i=0;
if (P&i>=10) then tpc&i=1;    /* Use (P&i>=8) for .08 Analyses          */
else tpc&i=0;
%end;
%mend mi;
%mi;
RUN;
```

(4) The next segment is the procedure **means** that computes the sum of **fpc&i,spc&i and tpc&i** and stores them in a dataset **case&i** for every imputation. Thus ten temporary datasets are created containing the sums **fsbac&i, ssbac&i and tsbac&i**. The dataset **drvbac** should first be sorted by **sex**.

```
proc sort data=drvbac;
by SEX;
run;

%macro DO_MEANS;
%do i=1 %to 10;
proc means noprint data=drvbac;
var fpc&i spc&i tpc&i;
by SEX;
output out=case&i n=total sum=fsbac&i ssbac&i tsbac&i;
run;
%end;
%mend DO_MEANS;
%DO_MEANS;
run;
```

(5) The next section of the code combines the ten datasets **case1**..**case10** and computes the Multiple-imputation point estimate $\overline{P}$ for each interval of study. For example, **pcnt0** is the $\overline{P}$ for the first interval of study, namely BAC=0.

```
data mi_est;
%macro AVG_EST;
%do i=1 %to 10;
set case&i;
sbac0=mean(fsbac1,fsbac2,fsbac3,fsbac4,fsbac5,fsbac6,fsbac7,fsbac8,fsbac9,fsb
ac10);
sbac1=mean(ssbac1,ssbac2,ssbac3,ssbac4,ssbac5,ssbac6,ssbac7,ssbac8,ssbac9,ssb
ac10);
sbac2=mean(tsbac1,tsbac2,tsbac3,tsbac4,tsbac5,tsbac6,tsbac7,tsbac8,tsbac9,tsb
ac10);
sbac3=sbac1+sbac2;
%end;
%mend AVG_EST;
%AVG_EST;
run;
```

(6) The next section of code tabulates the alcohol involvement percentages across the three categories and prints them to the output.

```
PROC TABULATE DATA=MI_EST FORMAT=COMMA10.0 MISSING;
WHERE SEX IN (1,2,9);
CLASS SEX;
VAR SBAC0 SBAC1 SBAC2 SBAC3 TOTAL;
TABLE SEX='Sex' ALL, (SBAC0='BAC=.00' SBAC1='BAC=.01-.09' SBAC2='BAC=.10+'
            TOTAL='Total Drivers Involved'
            SBAC3='Total Drivers w/Alcohol')*
            (SUM PCTSUM<TOTAL>) / RTS=15;
KEYLABEL N=' ' ALL='Total' SUM='Number' PCTSUM='Percent';
TITLE1 'DRIVERS INVOLVED IN FATAL CRASHES, BY SEX';
TITLE2 'FARS  1999';
RUN;
```

The routines tabulate an output as seen in Exhibit 3.

DRIVERS INVOLVED IN FATAL CRASHES, BY SEX
FARS 1999

	BAC=.00		BAC=.01-.09		BAC=.10+		Total Drivers Involved		Total Drivers w/Alcohol	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Sex										
Male	29,614	72	2,617	6	8,782	21	41,012	100	11,399	28
Female	12,720	86	528	4	1,588	11	14,835	100	2,116	14
Unknown	525	80	29	4	100	15	655	100	130	20
Total	42,858	76	3,174	6	10,470	19	56,502	100	13,644	24

Example 3: Program to determine 95% Confidence Interval and Standard Error of Estimate for the estimated Alcohol Involvement in Fatal Crashes

(1) Assign crash-level BAC into relevant categories. Of interest are the confidence intervals and standard error associated with the estimated alcohol involvement in fatal crashes, i.e., crashes where the crash BAC was 0.01 or greater.

```
DATA FIRST;
SET FARS&YR..ACC&YR;
%MACRO CATEGORIZE;
%DO I=1 %TO 10;
IF A&I=0 THEN FPC&I=1;          /*(BAC = 0.00)          */
ELSE FPC&I=0;
IF (1<=A&I<=9) THEN SPC&I=1;    /*(0.01<=BAC<=0.09)     */
ELSE SPC&I=0;
IF (A&I>=10) THEN TPC&I=1;      /*(BAC>=0.10)           */
ELSE TPC&I=0;
IF (A&I>=1) THEN ZPC&I=1;
ELSE ZPC&I=0;
%END;
%MEND CATEGORIZE;
%CATEGORIZE;
RUN;
```

(2) This part of the code computes the proportions for each imputation. Of interest is ZS_PP, which if the proportion of crashes for which the crash BAC is 0.01+.

```
%MACRO STATS;
%DO I=1 %TO 10;
PROC MEANS NOPRINT DATA=NULL1;
VAR FPC&I SPC&I TPC&I ZPC&I;
OUTPUT OUT=CASE&I N=TOTAL SUM=FSBAC&I SSBAC&I TSBAC&I ZSBAC&I
                    MEAN=FS_PP&I SS_PP&I TS_PP&I ZS_PP&I;
RUN;
%END;
%MEND STATS;
%STATS;
```

```
DATA CALC3;          /*     calculate variances for each imputation */
%MACRO VARIANCE;
%DO I=1 %TO 10;
SET CASE&I;
FS_VA&I=FS_PP&I*(1-FS_PP&I)/TOTAL;
SS_VA&I=SS_PP&I*(1-SS_PP&I)/TOTAL;
TS_VA&I=TS_PP&I*(1-TS_PP&I)/TOTAL;
ZS_VA&I=ZS_PP&I*(1-ZS_PP&I)/TOTAL;
%END;
%MEND VARIANCE;
%VARIANCE;
RUN;
```

(3) This section computes the parameters needed to evaluate the total variance of the estimates for the .01+ category. The variance estimate associated with the estimate has two components. The *within-imputation* variance is the average of the complete-data variance estimates,

$$\overline{U} = \frac{1}{10}\sum_{t=1}^{10} U^{(t)} \qquad (1)$$

and the *between-imputation* variance is the variance of the complete-data point estimates,

$$B = \frac{1}{9}\sum_{t=1}^{10} (\hat{Q}^{(t)} - \overline{Q})^2 \qquad (2)$$

The *total-variance* is defined as

$$T = \overline{U} + (1 + m^{-1})B \qquad (3)$$

where m is the number of imputations.

```
DATA P (KEEP=ZS_PP1-ZS_PP10 );    /*KEEPS THE TEN PROPORTIONS TO ESTIMATE TO    */
SET CALC3;                        /* ESTIMATE ‘B’ IN (2)     */
RUN;

PROC TRANSPOSE DATA=P OUT=PBAR;
RUN;

PROC MEANS DATA=PBAR noprint;     /*ESTIMATES ‘B’ AS IN (2)  */
VAR COL1;
OUTPUT OUT=PBAR_B MEAN=PBAR VAR=B;
RUN;
```

```
DATA U (KEEP=ZS_VA1-ZS_VA10 );    /*KEEPS THE TEN VARIANCES TO ESTIMATE */
SET CALC3;                        /* ESTIMATE Ū IN (1)      */
RUN;

PROC TRANSPOSE DATA=U OUT=UBAR;
RUN;

PROC MEANS DATA=UBAR noprint;     /*COMPUTES THE AVERAGE OF THE VARIANCES TO    */
VAR COL1;                         /*DERIVE Ū AS IN (1)
OUTPUT OUT=UBAR MEAN=UBAR;
RUN;
```

(4) This section combines the results according to Rubin (1987) to determine the total-variance and the 95% confidence intervals.

```
DATA STATS;
MERGE PBAR_B UBAR;
RUN;
```

The inferences are based on the approximation

$$T^{-1/2}(Q-\overline{Q}) \sim t_{n} \qquad (4)$$

where the degrees of freedom < are given by

$$n = (m-1)\left[1+\frac{\overline{U}}{(1+m^{-1})B}\right]^{2} \qquad (5)$$

Thus a 100(1-")% interval estimate for the estimate is given by

$$\overline{Q} \pm t_{n,1-a/2}\sqrt{T} \qquad (6)$$

The t distribution is evaluated in the code below using the TINV function whose arguments are < (NU) and 1-"/2 (1-.05/2 = 1-.025 = .975). The total variance TM is computed as in (1). The lower and upper bounds of the 95% confidence intervals are evaluated in LOW and HIGH, respectively in the code below.

```
DATA STATS;
SET STATS;
TM=UBAR+(1.1)*B;                                              /* T=U+(1+1/m)*B    */
RM=(1.1)*B/UBAR;
NU=9*(1+(1/RM))**2;
LOW=PBAR-TINV(.975,NU)*SQRT(TM);
HIGH=PBAR+TINV(.975,NU)*SQRT(TM);
RUN;

PROC TABULATE DATA=ALLSTATS out=test;
      VAR LOW PBAR HIGH SE;
      TABLE (LOW*F=5.4 PBAR='Estimate'*F=5.4 HIGH*F=5.4 SE='Std. Error'*f=10.8);
      TITLE1 ‘95% CONFIDENCE INTERVALS AND STANDARD ERROR OF ESTIMATE’;
      TITLE2 ‘ESTIMATE OF ALCOHOL RELATED CRASHES, FARS 1999’;
      RUN;
```

```
                95% CONFIDENCE INTERVALS AND STANDARD ERROR OF ESTIMATE
                     ESTIMATE OF ALCOHOL RELATED CRASHES, FARS 1999
```

LOW	Estimate	HIGH	Std. Error
Sum	Sum	Sum	Sum
.3922	.3977	.4032	0.00278626

LOW and HIGH are the lower and upper bounds of the 95 percent Confidence Interval associated with the estimate. The estimated proportion of crashes that are alcohol-related is .398 or 39.8%. The Standard error of estimate is 0.002786 or 0.2786%.