

METHODS USED TO IMPUTE ANNUAL FAMILY INCOME IN THE NATIONAL HEALTH INTERVIEW SURVEY, 1990-96

I. Introduction

There are two annual family income questions on the core questionnaire of the National Health Interview Survey (NHIS) for the years 1990-96. The first question asks whether family income is less than \$20,000 or \$20,000 or more while the second question asks participants to categorize their annual family income into one of 27 income groups. The wording for these two questions is given in table 1. The first question has approximately 3-4 percent of the values missing during 1990-96, while the second has a substantial amount of item missing data, approximately 16-18 percent (table 2).

Annual family income is an important measure of socioeconomic status and is strongly associated with health status, health insurance coverage, and access to health care. Thus, the annual family income questions are used for numerous analyses. The loss of 16-18 percent of sample data due to missing values on the detailed family income question raises questions about the extent to which results can be generalized to the entire population. Case-wise deletion is recognized to be problematic when a large proportion of cases are deleted from an analysis due to missing values. The concern is with the loss of precision due to deletion of cases and to the bias that may arise when cases with missing data are deleted. Thus, a decision was made to impute missing values for the detailed annual family income question, and for consistency, for the dichotomous annual family income question as well.

II. Selection of the imputation method

There are a variety of imputation techniques that could be used for the imputation of the two annual family income variables: cell mean imputation, hot deck imputation, flexible matching procedures, regression methods, regression methods calibrated to known or desired distributions for predictor variables, and multiple imputation. Choice among alternative imputation methods cannot be guided by clear theoretical reasons favoring one method over another. Empirical evidence is incomplete about which, if any, of the techniques is optimal for imputing a particular item or set of items. Thus, the choice of one technique over another is complex and often based on practical implementation issues.

When reviewing the alternative imputation procedures for use in the imputation of the two annual family income questions, the following criteria were considered:

Consistency with existing methods used on the NHIS data. The income and asset items in the NHIS 1990-1995 Family Resources Supplements were imputed using a sequential hot deck procedure within matrix cells. Analysts of NHIS data are familiar with this method. Employing a similar hot deck procedure would maintain consistency with past practice for similar items (income and assets), and reduce confusion among the user audience.

Statistical merit. The statistical literature does not provide strong evidence favoring one method over another. The sequential hot deck procedure, particularly when modified to incorporate the cell matrix feature of the cell mean approach, has been shown to perform reasonably well for individual item imputations relative to other methods (1).

Cost of programming and ease of implementation. Given that only two items were to be imputed, a method that could be readily programmed in an available programming language, such as SAS, was preferred.

These considerations led to the decision to use a sequential hot deck procedure within matrix cells. This approach was used for imputation in the NHIS Family Resources Supplements, has been shown to perform reasonably well for individual item imputations relative to other methods, and is relatively easy to program. In addition, hot deck imputation methods have been used to assign values to missing entries for income variables in various other national health surveys including: 1987 National Medical Expenditure Survey, Current Population Survey, Survey of Income and Program Participation, and National Survey of America's Families.

III. Sequential hot deck imputation within matrix cells

The sequential hot deck imputation within matrix cells is a combination of the cell mean imputation approach and the sequential hot deck imputation approach. Basically, the cell mean imputation procedure divides the sample into subgroups (cells) based on variables that are related to the item to be imputed (in this case, annual family income) and show variation in missing data rates across cells. The mean of the observed values within each cell is computed and used to replace missing values in the cell. Cell mean imputation has the methodological problem that the imputation does not have any stochastic element, creating clusters of cases in the data set all with the same value. Cell mean imputation can be modified to include stochastic variation by adding a randomly chosen residual from a specified distribution to the cell mean for each imputed value.

The goals of cell mean with stochastic residual imputation can also be achieved through a sequential hot deck procedure. Hot deck imputation begins with identification of a set of variables which are related to the item to be imputed. The data set is sorted with respect to these variables, ordering cases next to one another in the sorted list that are as alike as possible with respect to the sort variables. The mean of the observed values of the item to be imputed is computed and stored as an initial imputation value, referred to as a cold deck value. A simple sequential process then begins. The item value for the first case is examined, and if missing, is replaced with the cold deck value. If the value is non-missing, it replaces the cold deck value, becoming the hot deck value. The sequential process continues case-by-case through the data set until all cases have been processed. Item missing values are thus replaced by the last nearest neighbor in the list that had a non-missing value.

The sequential hot deck procedure may tend to use the same hot deck value from the same donor repeatedly, and may make imputations across class boundaries of sort variables that are not

desirable. The method can be altered to a cell matrix sequential hot deck in which the data set is first divided into matrix cells based on a cross-classification of key variables related to the imputation item, and the sequential hot deck process is repeated separately in each matrix cell. Separate cold deck values are needed for each matrix cell.

IV. Implementation of the hot deck imputation procedure

A separate imputation was performed for each of the survey years 1990-96. For the 1993 survey year, two separate imputations were performed, one for the portion of the sample that did not receive the Family Resources Supplement (January - June 1993) and one for the portion that did receive the Supplement (July - December 1993). For the 1996 survey year, the Family Resources Supplement had not been imputed at the time of this project and so data from the supplement were not used in the imputation for 1996. For consistency, the same matrix cross-classification (stratification) and sort variables were used for the imputations performed for 1990-92, July - December 1993, 1994 and 1995. The cross-classification and sort variables used for January -June, 1993 and for 1996 were the same as each other but differed from those used for the other survey years.

The dichotomous family income variable (less than \$20,000, \$20,000 and over), which had a missing rate of about 3-4 percent, was imputed first. The detailed (27 category) family income variable, which had a missing rate of about 16-18 percent, was imputed second. The imputation of the detailed annual family income variable was performed separately for families with less than \$20,000 income and those with at least \$20,000 income, based on the value of the dichotomous annual family income variable (the imputed values for this variable were used if necessary). This was done so that information from the dichotomous annual family income variable, which had less missing data, could be used in the imputation of the detailed annual family income variable and so that the imputed detailed income value would be consistent with the dichotomous income value.

A. Creation of family-level files

For each survey year a family-level file was created that included one record for each family with selected segment variables (e.g. median annual family income for the families within a segment), family variables (e.g. reported annual family income) and family reference person variables (e.g. age of family's reference person). The reference person is the member of the family who owns or rents the dwelling unit). The imputation for each survey year was performed using a family-level file so that all members of the same family would be assigned the same family income value. For 1990-95, the number of families ranged from 40,236 in 1995 to 50,556 in 1992. The number of families in 1996 was 24,980.

A small number of families had no adult member (no person with age 18 years or over). For these families, if there was a person 16 or 17 years of age, that person's values were used as the reference person values for the imputation and the family's reference person was assigned to the 18-24 year age group for subsequent analyses.

For a small number of families, not all members of the family had the same value for the dichotomous family income variable due to an editing error. For these families, the dichotomous income variable was edited prior to the imputation as follows: (1) if there were two family members, and they had differing non-missing income values, the reference person's income value was assigned to the other family member; (2) if there were three or more family members and they had differing, non-missing income values, the majority income value was assigned to all family members, (3) if there were three or more family members and one or more of them had a missing income value, the majority non-missing income value was assigned to those family members.

Segment income level was considered as a possible predictor of annual family income as families from the same sample segment are neighbors and hence tend to have similar family incomes. Two segment income variables were created for each family-level file: (1) mean of reported dichotomous annual family income values within each sample segment and (2) median of reported detailed annual family income values within each sample segment.

(1) The mean of reported dichotomous annual family income values for families residing within a sample segment was calculated within a sample segment. For a small number of segments, no families within the segment reported annual income for the dichotomous family income question. The mean segment family income for these segments was imputed.

(2) The median of reported detailed family income values for families residing within a sample segment was calculated for each segment. For some segments, no families within the segment reported annual income on the detailed income question. The median segment income for these segments was imputed.

Occupational pay category for the family reference person was obtained by recoding the 15-category Occupation Recode No. 2 variable from the core NHIS into four groups: low-paying occupations, medium-paying occupations, high-paying occupations, and not in labor force. This 4-category recode was used in the imputation of missing values for the Family Resource Supplement.

B. Selection of cross-classification and sort variables

The first step in the sequential hot deck imputation within matrix cells procedure was to select the variables to be used to form the matrix cells and the variables to be used to sort the records within each cell. The variables that were the strongest predictors of annual family income were selected for use as the matrix cross-classification and sort variables. Variables considered as potential cross-classification and sort variables included those used in the imputation of personal monthly income and assets on the NHIS Family Resources Supplement, as well as variables from the NHIS Family Resources Supplement. As the imputation of the annual family income variables was done using family-level files, the variables considered for use in the imputation either provided information about family characteristics or about the family's reference person.

The imputations for both the dichotomous and detailed annual family income variables were performed separately for the following three groups based on the age of the family's reference person: (1) 18-24 years of age, (2) 25-64 years of age, and (3) 65 years of age and over. This was done because families with reference persons in each of these three age groups are likely to have different levels and sources of income as well as potentially different predictors of annual family income. Those under age 25 may not have completed their education. Those 25-64 years of age are likely to have completed their education and be employed, and those 65 and over may be retired. In addition the percent of persons with missing annual family income was higher for persons age 65 years and over than for younger persons.

Dichotomous annual family income imputation

Univariate logistic regressions were performed to identify significant predictors of dichotomous annual family income. The dichotomous income variable was the dependent variable and selected family-level and family reference person variables were the independent variables.

Stepwise logistic regression models with the variables found to be significant predictors of annual family income were used to select the cross-classification and sort variables (see table 3). Family income in the month prior to interview from the NHIS Family Resources Supplement and mean segment income were categorized into age group-specific quartiles.

Those variables that changed the likelihood ratio the most were used to form the matrix cells. If the use of a variable to form the cells resulted in a large number of cells with fewer than 25 families or with a ratio of donors to missing of less than 2, that variable was not used for the matrix cross-classification. Because the number of families in the three age groups differs, the 18-24 year age group being the smallest and the 25-64 year age group being the largest, the number of variables that could be used to define the matrix cells differed by age group. Once the cross-classification variables were selected and the matrix cells defined, any cells with fewer than 25 families or with a donors to missing ratio less than 2 were combined with another cell. Variables that entered into the stepwise regression models but were not used as cross-classification variables were used as sort variables. In addition, a random component was added to the sort list so that families with the same sort values would be randomly ordered.

The cross-classification and sort variables for 1990-92, 1993(July-December), 1994, and 1995 were identified using regressions that were performed with data for 1993(July-December, 1994 and 1995 pooled). The cross-classification and sort variables for 1993(January-June) and 1996 were identified using regressions that were performed with data pooled for 1993(January-June) and 1996. See table 4 for a complete list of the cross-classification and sort variables for each age group.

Detailed annual family income imputation

Univariate ordinary least squares regressions were performed to identify which variables were significant predictors of detailed annual family income. Regressions were performed separately for income codes 0-19 (annual family income less than \$20,000) and for the income codes 20-26 (annual family income \$20,000 and over). The regressions were done in two sets because for codes 0-19, the income increment is \$1,000 whereas for codes 20-26 the income increment is \$5,000 (with the cap of \$50,000 and over for the final code of 26). Also, this break is in keeping with the stratification on income of less than \$20,000 and \$20,000 and over.

Stepwise regression models with the variables found to be significant predictors of annual family income were used to select the cross-classification and sort variables (see table 3). Family income in the month prior to interview from the NHIS Family Resources Supplement and median segment income were categorized into income group-specific (less than \$20,000 and \$20,000 and over) and age group-specific quartiles. Income group-age group-specific quartiles were used because the monthly family income and median segment income distributions differed substantially by income group and age group.

Variables that increased the R-square by 1 percent or more were selected for use as matrix cross-classification variables. If the use of a variable to form the matrix cells resulted in a large number of cells with fewer than 25 families or with a ratio of donors to missing of less than 2, that variable was not used for the cross-classification. Because of the much larger percent of missing data for the detailed income variable, the ratio of donors to missing records was less than 2 more often than for the dichotomous variable and thus, fewer cross-classification variables could be used. Once the cross-classification variables were selected and the matrix cells defined, any cells with fewer than 25 families or with a donor to missing ratio less than 2 were combined with another cell. The variables that entered into the stepwise regression models but were not used as cross-classification variables were used as sort variables. In addition, a random component was added to the sort list so that families with the same sort values would be randomly ordered.

The cross-classification and sort variables for 1990-92, 1993(July-December), 1994, and 1995 were identified using regressions that were performed with the data for 1993(July-December, 1994 and 1995 pooled). The cross-classification and sort variables for 1993(January-June) and 1996 were identified using regressions that were performed with the data for 1993(January-June) and 1996 pooled. See table 5 for a complete list of cross-classification and sort variables for each age-income group

V. Results of the hot deck imputation

The results of the dichotomous and detailed annual family income hot deck imputations are shown in Tables 6-12. For each of the seven survey years, the percent of persons with an imputed income of less than \$20,000 was greater than the percent who reported an income of less than \$20,000. The imputation of the detailed income variable was contingent upon the imputation of the dichotomous variable. Thus, the larger percentage of persons with imputed incomes less than \$20,000 than with reported incomes of less than \$20,000 that was obtained in the dichotomous imputation resulted in larger percentages of individuals with imputed incomes in the 20 income categories between \$0 and \$20,000 relative to the percentages with those reported values. The \$20,000-\$24,999 category generally had a slightly larger percent of imputed values assigned to it whereas the categories between \$25,000 and \$49,999 were slightly reduced. Considerably fewer persons were assigned imputed values of \$50,000 and over than reported this income range.

VI. Calculation of variances of imputed variables

Calculation of the variances of imputed variables is an important issue. To calculate the variance of an imputed variable correctly, the analyst must account for the sampling variation of both the observed values and the imputed values. If the imputed values are treated as observed values, the resulting variance may substantially underestimate the true variance.

There are several alternative methods for variance estimation when using imputed values. One approach employs jackknife variance estimation methods (2). Under this approach, the variance of an estimate can be properly computed using adjusted values across jackknife replicates of the sample. The method is not yet readily available in existing software. There are also several “model-assisted” methods for obtaining variance estimates which are applicable to hot deck imputation (3). Although these methods exist and have been applied to survey data with imputed values, routine application for most analysts of survey data is not straightforward.

Other methods for obtaining correct variances when using imputed values are not appropriate for the hot deck imputation procedure used for the NHIS annual family income variables (4-5). For example, under multiple imputation, the variance of the imputation process can be estimated directly from a computation of the variance between and within multiply imputed values (4).

Analysts using the imputed annual family income variables for 1990-96 should be aware that methods exist for properly accounting for imputation variance in estimates, but until practical software implementation of these methods occurs, should not expect to apply these methods routinely to NHIS estimation.

References

1. Nordholt ES. Imputation: methods, simulation experiments and practical examples. *International Statistical Review* 66 (2): 157-180, 1998.
2. Rao, JNK, and Shao J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* 79: 811-822, 1992.
3. Sarndal CE. Methods for estimating the precision of survey estimates when imputation has been used. *Survey Methodology* 18 (2): 241-252, 1992.
4. Rubin DR. *Multiple Imputation for Nonresponse in Surveys*. New York: J.W. Wiley and Sons, Inc. 1987.
5. Fay, RE. Valid inferences from imputed survey data. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 227-232, 1993.

Table 1. NHIS Annual Family Income Questions

Question	Response
<p>Was the total combined FAMILY income during the past 12 months-that is, yours (<i>read names, including Armed Forces members living at home</i>) more or less than \$20,000? Include money from jobs, social security, retirement income, unemployment payments, public assistance, and so forth. Also include income from interest, dividends, net income from business, farm, or rent, and any other money income received.</p>	<p>1. \$20,000 or more 2. Less than \$20,000</p>
<p>Of those income groups, which letter best represents the total combined FAMILY income during the past 12 months that is, yours (<i>read names, including Armed Forces members living at home</i>)? Include wages, salaries, and other items we just talked about.</p>	<p>A. Less than \$1,000 (including loss) B. \$1,000-\$1,999 C. \$2,000-\$2,999 D. \$3,000-\$3,999 E. \$4,000-\$4,999 F. \$5,000-\$5,999 G. \$6,000-\$6,999 H. \$7,000-\$7,999 I. \$8,000-\$8,999 J. \$9,000-\$9,999 K. \$10,000-\$10,999 L. \$11,000-\$11,999 M. \$12,000-\$12,999 N. \$13,000-\$13,999 O. \$14,000-\$14,999 P. \$15,000-\$15,999 Q. \$16,000-\$16,999 R. \$17,000-\$17,999 S. \$18,000-\$18,999 T. \$19,000-\$19,999 U. \$20,000-\$24,999 V. \$25,000-\$29,999 W. \$30,000-\$34,999 X. \$35,000-\$39,999 Y. \$40,000-\$44,999 Z. \$45,000-\$49,999 ZZ. \$50,000 and over</p>

Table 2. Percent of persons with missing values for the dichotomous and detailed annual family income questions in the National Health Interview Survey, 1990-96

Survey year	Percent missing	
	Dichotomous	Detailed
	question	question
1990	2.9	16.7
1991	3.3	18.1
1992	3.3	18.4
1993	3.2	16.1
1994	3.8	17.2
1995	3.9	16.3
1996	4.1	17.1

Table 3. Variables included in the stepwise multivariate regressions used to select the cross-classification and sort variables for the imputation.

Sample segment variables

Segment income	Segment income refers to the mean or median annual family income for residents of a sample segment. Families from the same segment are neighbors and hence tend to have similar family incomes. These variables were categorized into quartiles.
----------------	--

Family variables

Monthly family income	Monthly family income in the month prior to interview is a continuous variable from the Family Resources Supplement of the NHIS obtained by summing the personal monthly income of each family member. Values for personal monthly income were previously imputed for all persons with missing data.
-----------------------	--

No. of adult workers	(0, 1, 2 or more)
----------------------	-------------------

Family reference person variables

Age	18-24 years, 25-64 years, 65 years and over
-----	---

Sex	male, female
-----	--------------

Race-ethnicity	non-Hispanic black, Hispanic, all other
----------------	---

Marital status	married, not married
----------------	----------------------

Educational attainment	less than high-school, high-school, some college, college graduate or more
------------------------	--

Occupational pay category	low-paying, mid-paying, high-paying jobs, not in labor force
---------------------------	--

Usual hours worked per week	less than 35 hours, 35 hours or more
-----------------------------	--------------------------------------

Table 4. Cross-classification and sort variables used for the dichotomous annual family income imputation by age group and survey year

Survey year and age group	Cross-classification variables	Sort variables
1990, 1991, 1992, 1993 (July-December), 1994, 1995		
18-24 years	Monthly family income Mean segment income	No. of adult workers in family Occupational pay category Marital status Educational attainment
25-64 years	Monthly family income Mean segment income Usual hours worked per week	Educational attainment Occupational pay category
65 years and over	Monthly family income Mean segment income Marital status	Educational attainment No. adult workers in family Occupational pay category Sex
1993 (January-June), 1996		
18-24 years	Mean segment income No. adult workers in family	Educational attainment Marital status Occupational pay category
25-64 years	Mean segment income No. adult workers in family Occupational pay category	Marital status Educational attainment Race-ethnicity Sex
65 years and over	Mean segment income No. adult workers in family	Educational attainment Sex Marital status

Note: See table for variable definitions.

Table 5. Cross-classification and sort variables used for the detailed annual family income imputation by age-income group and survey year.

Survey year and age-income group	Cross-classification variables	Sort variables
1990, 1991, 1992, 1993 (July-December), 1994, 1995		
18-24 years less than \$20,000	Monthly family income Median segment income	Marital status Usual hours worked per week No. adult workers in family Occupational pay category
\$20,000 and over	Monthly family income Median segment income	Educational attainment Usual hours worked per week
25-64 years less than \$20,000	Monthly family income Median segment income Usual hours worked per week	Marital status No. adult workers in family Occupational pay category Educational attainment
\$20,000 and over	Monthly family income Median segment income Marital status	Educational attainment Occupational pay category Sex
65 years and over less than \$20,000	Monthly family income Median segment income	Marital status Educational attainment Sex Race-ethnicity
\$20,000 and over	Monthly family income Median segment income	Educational attainment Usual hours worked per week Marital status No. adult workers in family

Table 5 (continued)

Survey year and age-income group	Cross-classification variables	Sort variables
1993 (January-June), 1996		
18-24 years less than \$20,000	Median segment income No. adult workers in family	Marital status Sex Educational attainment
\$20,000 and over	Median segment income	Occupational pay category No. adult workers in family Marital status
25-64 years less than \$20,000	Median segment income Occupational pay category Marital status	No. adult workers in family Educational attainment
\$20,000 and over	Median segment income No. adult workers in family Educational attainment	Marital status Occupational pay category Sex
65 years and over less than \$20,000	Median segment income Marital status	Educational attainment No. adult workers in family Sex
\$20,000 and over	Median segment income Educational attainment	No. adult workers in family Marital status Occupational pay category

Note: See table 3 for variable definitions.

Table 6. Distribution of 1990 annual family income, by imputation status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	39,525	32.81	1,427	39.67	40,952	33.01
\$20,000 and over	76,632	67.19	2,047	60.33	78,679	66.99
Less than \$1,000	596	0.59	141	0.66	737	0.60
\$1,000-1,999	722	0.70	166	0.75	888	0.71
\$2,000-2,999	833	0.80	137	0.64	970	0.78
\$3,000-3,999	984	0.93	220	1.02	1,204	0.94
\$4,000-4,999	1,425	1.33	309	1.53	1,734	1.36
\$5,000-5,999	1,441	1.33	335	1.58	1,776	1.37
\$6,000-6,999	1,635	1.54	477	2.29	2,112	1.66
\$7,000-7,999	1,578	1.48	399	2.00	1,977	1.56
\$8,000-8,999	1,577	1.54	411	1.97	1,988	1.61
\$9,000-9,999	1,666	1.58	393	1.88	2,059	1.63
\$10,000-10,999	2,342	2.26	563	2.69	2,905	2.33
\$11,000-11,999	1,511	1.49	478	2.16	1,989	1.60
\$12,000-12,999	2,490	2.43	656	3.21	3,146	2.56
\$13,000-13,999	1,539	1.50	375	1.82	1,914	1.55
\$14,000-14,999	1,529	1.50	456	2.31	1,985	1.64
\$15,000-15,999	2,099	2.06	633	3.07	2,732	2.22
\$16,000-16,999	1,709	1.65	481	2.30	2,190	1.76
\$17,000-17,999	1,676	1.65	554	2.65	2,230	1.81
\$18,000-18,999	2,174	2.12	682	3.35	2,856	2.32
\$19,000-19,999	2,642	2.69	918	4.65	3,560	3.01
\$20,000-24,999	9,036	8.88	2,096	10.37	11,132	9.12
\$25,000-29,999	8,375	8.52	1,699	8.87	10,074	8.58
\$30,000-34,999	8,321	8.40	1,612	8.26	9,933	8.38
\$35,000-39,999	6,842	6.97	1,237	6.26	8,079	6.86
\$40,000-44,999	6,506	6.66	1,058	5.38	7,564	6.45
\$45,000-49,999	5,539	5.75	867	4.47	6,406	5.54
\$50,000 and over	22,872	23.68	2,619	13.88	25,491	22.07

Table 7. Distribution of 1991 annual family income, by imputation status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	38,164	31.90	1,686	41.29	39,850	32.21
\$20,000 and over	77,868	68.10	2,314	58.71	80,182	67.79
Less than \$1,000	613	0.65	165	0.75	778	0.67
\$1,000-1,999	680	0.67	174	0.76	854	0.69
\$2,000-2,999	732	0.71	125	0.58	857	0.68
\$3,000-3,999	966	0.91	249	1.10	1,215	0.94
\$4,000-4,999	1,301	1.24	338	1.49	1,639	1.28
\$5,000-5,999	1,406	1.37	447	1.97	1,853	1.47
\$6,000-6,999	1,620	1.52	478	2.13	2,098	1.63
\$7,000-7,999	1,623	1.56	474	2.05	2,097	1.65
\$8,000-8,999	1,548	1.47	418	1.78	1,966	1.53
\$9,000-9,999	1,670	1.60	490	2.14	2,160	1.70
\$10,000-10,999	1,945	1.91	587	2.66	2,532	2.04
\$11,000-11,999	1,420	1.42	406	1.91	1,826	1.51
\$12,000-12,999	2,497	2.53	812	3.70	3,309	2.74
\$13,000-13,999	1,491	1.53	408	1.79	1,899	1.58
\$14,000-14,999	1,574	1.56	411	1.80	1,985	1.60
\$15,000-15,999	2,054	2.06	616	2.90	2,670	2.21
\$16,000-16,999	1,554	1.53	444	2.06	1,998	1.63
\$17,000-17,999	1,529	1.54	513	2.37	2,042	1.68
\$18,000-18,999	2,125	2.10	673	3.16	2,798	2.29
\$19,000-19,999	2,429	2.42	845	3.89	3,274	2.68
\$20,000-24,999	9,146	9.22	2,385	11.15	11,531	9.56
\$25,000-29,999	7,995	8.10	1,729	7.85	9,724	8.05
\$30,000-34,999	8,448	8.72	1,823	8.38	10,271	8.66
\$35,000-39,999	6,677	6.85	1,425	6.66	8,102	6.81
\$40,000-44,999	6,460	6.64	1,186	5.72	7,646	6.48
\$45,000-49,999	5,559	5.71	1,080	5.05	6,639	5.59
\$50,000 and over	23,287	24.47	2,982	14.20	26,269	22.64

Table 8. Distribution of 1992 annual family income, by imputation status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	42,419	32.52	1,651	38.66	44,070	32.72
\$20,000 and over	81,838	67.48	2,504	61.34	84,342	67.28
Less than \$1,000	642	0.64	172	0.73	814	0.66
\$1,000-1,999	762	0.68	181	0.64	943	0.67
\$2,000-2,999	853	0.77	179	0.74	1,032	0.77
\$3,000-3,999	966	0.90	236	0.96	1,202	0.91
\$4,000-4,999	1,322	1.23	318	1.25	1,640	1.23
\$5,000-5,999	1,634	1.43	544	2.16	2,178	1.56
\$6,000-6,999	1,768	1.59	613	2.42	2,381	1.74
\$7,000-7,999	1,766	1.55	520	2.07	2,286	1.65
\$8,000-8,999	1,740	1.57	578	2.31	2,318	1.70
\$9,000-9,999	1,791	1.60	553	2.08	2,344	1.68
\$10,000-10,999	2,502	2.29	750	3.20	3,252	2.46
\$11,000-11,999	1,464	1.33	440	1.81	1,904	1.42
\$12,000-12,999	2,657	2.41	822	3.26	3,479	2.56
\$13,000-13,999	1,749	1.58	538	2.09	2,287	1.68
\$14,000-14,999	1,644	1.51	490	2.01	2,134	1.60
\$15,000-15,999	2,354	2.12	766	2.98	3,120	2.27
\$16,000-16,999	1,617	1.50	589	2.28	2,206	1.64
\$17,000-17,999	1,484	1.36	493	1.97	1,977	1.47
\$18,000-18,999	2,213	2.08	698	2.95	2,911	2.24
\$19,000-19,999	2,702	2.53	960	4.00	3,662	2.80
\$20,000-24,999	9,447	8.83	2,436	10.55	11,883	9.15
\$25,000-29,999	8,303	8.00	1,932	8.40	10,235	8.07
\$30,000-34,999	8,549	8.13	1,793	7.67	10,342	8.05
\$35,000-39,999	6,889	6.65	1,411	6.26	8,300	6.58
\$40,000-44,999	6,802	6.65	1,214	5.45	8,016	6.43
\$45,000-49,999	5,907	5.82	963	4.20	6,870	5.52
\$50,000 and over	25,218	25.28	3,478	15.54	28,696	23.49

Table 9. Distribution of 1993 annual family income, by imputation status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	33,967	31.17	1,180	33.11	35,147	31.24
\$20,000 and over	72,201	68.83	2,323	66.89	74,524	68.76
Less than \$1,000	576	0.64	126	0.80	702	0.67
\$1,000-1,999	619	0.65	101	0.49	720	0.63
\$2,000-2,999	775	0.81	162	0.93	937	0.83
\$3,000-3,999	813	0.82	186	0.94	999	0.84
\$4,000-4,999	1,017	1.06	176	0.93	1,193	1.04
\$5,000-5,999	1,360	1.37	355	1.87	1,715	1.45
\$6,000-6,999	1,408	1.46	385	2.10	1,793	1.57
\$7,000-7,999	1,398	1.47	317	1.76	1,715	1.52
\$8,000-8,999	1,386	1.45	343	1.95	1,729	1.53
\$9,000-9,999	1,750	1.78	484	2.66	2,234	1.92
\$10,000-10,999	1,994	2.09	481	2.68	2,475	2.18
\$11,000-11,999	1,208	1.31	341	1.81	1,549	1.39
\$12,000-12,999	2,128	2.26	549	3.04	2,677	2.38
\$13,000-13,999	1,306	1.39	352	1.99	1,658	1.48
\$14,000-14,999	1,387	1.51	407	2.30	1,794	1.63
\$15,000-15,999	1,793	1.96	564	3.27	2,357	2.17
\$16,000-16,999	1,300	1.39	393	2.20	1,693	1.52
\$17,000-17,999	1,447	1.55	378	2.21	1,825	1.66
\$18,000-18,999	1,871	1.99	485	2.64	2,356	2.10
\$19,000-19,999	2,358	2.53	668	3.89	3,026	2.74
\$20,000-24,999	8,301	9.04	1,873	10.56	10,174	9.28
\$25,000-29,999	6,998	7.68	1,352	7.77	8,350	7.70
\$30,000-34,999	7,746	8.50	1,316	7.62	9,062	8.36
\$35,000-39,999	5,992	6.53	998	5.54	6,990	6.37
\$40,000-44,999	6,052	6.58	995	5.86	7,047	6.47
\$45,000-49,999	5,486	6.08	966	5.49	6,452	5.98
\$50,000 and over	23,574	26.11	2,875	16.68	26,449	24.60

Table 10. Distribution of 1994 annual family income, by imputation status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	34,351	29.86	1,545	33.51	35,896	30.00
\$20,000 and over	77,463	70.14	2,820	66.49	80,283	70.00
Less than \$1,000	577	0.61	167	0.79	744	0.64
\$1,000-1,999	696	0.69	123	0.61	819	0.67
\$2,000-2,999	715	0.74	179	0.90	894	0.77
\$3,000-3,999	759	0.79	151	0.77	910	0.79
\$4,000-4,999	916	0.89	251	1.13	1,167	0.93
\$5,000-5,999	1,311	1.31	402	1.78	1,713	1.39
\$6,000-6,999	1,311	1.28	395	1.81	1,706	1.37
\$7,000-7,999	1,548	1.50	383	1.94	1,931	1.57
\$8,000-8,999	1,399	1.40	346	1.69	1,745	1.45
\$9,000-9,999	1,671	1.66	452	2.17	2,123	1.75
\$10,000-10,999	2,083	2.05	613	3.00	2,696	2.21
\$11,000-11,999	1,185	1.20	357	1.75	1,542	1.30
\$12,000-12,999	2,141	2.18	600	2.95	2,741	2.31
\$13,000-13,999	1,367	1.41	361	1.78	1,728	1.47
\$14,000-14,999	1,439	1.44	440	2.17	1,879	1.56
\$15,000-15,999	1,917	2.01	557	2.67	2,474	2.12
\$16,000-16,999	1,346	1.34	394	1.97	1,740	1.45
\$17,000-17,999	1,428	1.44	475	2.41	1,903	1.61
\$18,000-18,999	1,807	1.91	537	2.57	2,344	2.02
\$19,000-19,999	2,301	2.36	796	3.88	3,097	2.62
\$20,000-24,999	8,804	9.11	2,144	10.91	10,948	9.42
\$25,000-29,999	7,413	7.74	1,488	7.66	8,901	7.72
\$30,000-34,999	7,866	8.33	1,658	8.41	9,524	8.35
\$35,000-39,999	6,446	6.73	1,152	5.71	7,598	6.56
\$40,000-44,999	6,412	6.74	1,221	6.13	7,633	6.64
\$45,000-49,999	5,533	5.81	915	4.65	6,448	5.61
\$50,000 and over	25,766	27.35	3,465	17.81	29,231	25.71

Table 11. Distribution of 1995 Annual Family Income, by Imputation Status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	31,016	28.40	1,455	34.60	32,471	28.64
\$20,000 and over	67,486	71.60	2,510	65.40	69,996	71.36
Less than \$1,000	578	0.65	178	0.97	756	0.70
\$1,000-1,999	559	0.58	109	0.62	668	0.59
\$2,000-2,999	587	0.58	114	0.75	701	0.61
\$3,000-3,999	681	0.68	214	1.14	895	0.75
\$4,000-4,999	768	0.80	157	0.82	925	0.81
\$5,000-5,999	1,090	1.13	339	1.89	1,429	1.25
\$6,000-6,999	1,255	1.27	301	1.70	1,556	1.34
\$7,000-7,999	1,277	1.28	321	1.84	1,598	1.37
\$8,000-8,999	1,309	1.30	285	1.70	1,594	1.36
\$9,000-9,999	1,454	1.52	347	1.94	1,801	1.58
\$10,000-10,999	1,945	2.00	450	2.58	2,395	2.09
\$11,000-11,999	1,285	1.34	305	1.81	1,590	1.41
\$12,000-12,999	2,146	2.20	601	3.01	2,747	2.33
\$13,000-13,999	1,272	1.40	366	2.01	1,638	1.49
\$14,000-14,999	1,455	1.49	412	2.14	1,867	1.59
\$15,000-15,999	1,767	1.97	474	2.64	2,241	2.07
\$16,000-16,999	1,300	1.46	448	2.27	1,748	1.58
\$17,000-17,999	1,178	1.25	277	1.67	1,455	1.32
\$18,000-18,999	1,710	1.79	519	2.80	2,229	1.95
\$19,000-19,999	2,025	2.23	613	3.65	2,638	2.45
\$20,000-24,999	7,879	8.76	1,711	10.34	9,590	9.01
\$25,000-29,999	6,739	7.81	1,533	9.15	8,272	8.02
\$30,000-34,999	6,823	8.13	1,196	7.46	8,019	8.02
\$35,000-39,999	5,436	6.62	989	6.17	6,425	6.55
\$40,000-44,999	5,367	6.56	907	5.68	6,274	6.42
\$45,000-49,999	4,904	6.10	802	5.07	5,706	5.94
\$50,000 and over	22,944	29.09	2,766	18.19	25,710	27.40

Table 12. Distribution of 1996 Annual Family Income, by Imputation Status

Family Income	Reported		Imputed		Total	
	Sample Size	Weighted Percent	Sample Size	Weighted Percent	Sample Size	Weighted Percent
Less than \$20,000	18,635	27.21	815	29.81	19,450	27.31
\$20,000 and over	42,186	72.79	1,766	70.19	43,952	72.69
Less than \$1,000	243	0.46	102	0.84	345	0.52
\$1,000-1,999	366	0.64	69	0.50	435	0.62
\$2,000-2,999	319	0.54	59	0.43	378	0.52
\$3,000-3,999	356	0.61	91	0.74	447	0.63
\$4,000-4,999	433	0.70	117	0.94	550	0.74
\$5,000-5,999	694	1.13	206	1.49	900	1.19
\$6,000-6,999	695	1.16	198	1.43	893	1.21
\$7,000-7,999	699	1.12	140	1.27	839	1.14
\$8,000-8,999	701	1.17	233	1.99	934	1.31
\$9,000-9,999	941	1.48	229	1.82	1,170	1.54
\$10,000-10,999	1,132	1.88	337	2.77	1,469	2.03
\$11,000-11,999	705	1.16	213	1.69	918	1.25
\$12,000-12,999	1,304	2.21	348	3.04	1,652	2.35
\$13,000-13,999	838	1.41	241	1.86	1,079	1.48
\$14,000-14,999	828	1.39	186	1.60	1,014	1.43
\$15,000-15,999	1,100	1.96	341	2.96	1,441	2.12
\$16,000-16,999	777	1.28	171	1.47	948	1.31
\$17,000-17,999	810	1.45	254	2.20	1,064	1.58
\$18,000-18,999	1,042	1.81	339	2.83	1,381	1.98
\$19,000-19,999	1,216	2.17	377	3.36	1,593	2.37
\$20,000-24,999	4,786	8.63	1,123	10.73	5,909	8.98
\$25,000-29,999	4,130	7.95	753	7.15	4,883	7.82
\$30,000-34,999	4,150	7.96	743	7.18	4,893	7.83
\$35,000-39,999	3,179	6.30	704	6.59	3,883	6.35
\$40,000-44,999	3,180	6.32	562	5.65	3,742	6.21
\$45,000-49,999	2,824	5.71	549	5.47	3,373	5.67
\$50,000 and over	15,137	31.39	2,132	21.99	17,269	29.83