
NATIONAL CENTER FOR EDUCATION STATISTICS

User's Manual

July 1997

National Household Education Survey of 1996

**Data File User's Manual
Volume V**

Adult Civic Involvement Data File



NATIONAL CENTER FOR EDUCATION STATISTICS

User's Manual

July 1997

National Household Education Survey of 1996

Data File User's Manual Volume V

Adult Civic Involvement Data File



Mary A. Collins, Project Director
J. Michael Brick, Senior Statistician
Mary Jo Nolin
Susan Gilmore
Westat, Inc.

Kathryn Chandler
National Center for Education Statistics

Chris Chapman
Education Statistics Services Institute

U.S. Department of Education
Office of Educational Research and Improvement

NCES 97-421

U.S. Department of Education

Richard W. Riley
Secretary

Office of Educational Research and Improvement

Ramon C. Cortines
Acting Assistant Secretary

National Center for Education Statistics

Pascal D. Forgione, Jr.
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208-5574

July 1997

The NCES World Wide Web Home Page is
<http://www.ed.gov/NCES/>

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *National Household Education Survey of 1996: Data File User's Manual, Volume V: Adult Civic Involvement Data File*, NCES 97-421, by Mary Collins, J. Michael Brick, Mary Jo Nolin, Susan Gilmore, Kathryn Chandler and Chris Chapman. Washington, DC: 1997.

Contact:

Kathryn Chandler
(202) 219-1767
(e-mail) nhes@ed.gov
<http://www.ed.gov/NCES/NHES>

CONTENTS

Chapter

Introduction	1
6. Guide to the Data File and Codebook	3
6.1 Content and Organization of the Data File	3
6.1.1 System Variables.....	3
6.1.2 Household Membership Variables	4
6.1.3 Questionnaire Item Variables	5
6.1.4 Household Characteristics Variables.....	5
6.1.5 Derived Variables	5
6.1.6 Weighting and Variance Estimation Variables	8
6.1.7 Imputation Flag Variables	8
6.1.8 Numeric and Character Variables	8
6.2 Guide to the Codebook.....	8
6.3 Public and Proprietary Data Files	9
6.4 Linking the Household & Library File to Other NHES:96 Data Files	9
7. Data Considerations and Anomalies	11
7.1 Type of Community in Which the Household is Located	11
7.2 Income to the Nearest Thousand Dollars	11

Appendixes

Appendix B: Adult CI Public File Layout in Position Order.....	B-1
Appendix C: SAS Code for Derived Variables.....	C-1
Appendix D: Adult CI Codebook.....	D-1
Appendix E: Directions and Code for Linking Data Files	E-1

Figure

6-1 Example of the codebook format.....	10
---	----

This page is blank.

INTRODUCTION

The 1996 National Household Education Survey (NHES:96) was a random digit dial (RDD) telephone survey of households developed by the National Center for Education Statistics (NCES) and conducted by Westat, Inc. The NHES:96 included two topical survey components:

- Parent and Family Involvement in Education (PFI), in which data were collected about types and frequency of family involvement in children's school, school practices to involve and support families, and learning activities with children outside of school; and
- Civic Involvement (CI), which included sources of information about government, knowledge about government, community service participation, political participation, and attitudes related to democratic values and government.

There were three populations of interest for the NHES:96:

- Children 3 years old through grade 12, whose parents responded to PFI items, and children in 6th through 12th grades, whose parents also responded to CI items;
- Students in grades 6 through 12, who, in addition to their parents, responded to CI items and to a small number of PFI items; and
- Adults, defined as persons 18 years old or older, not enrolled in grade 12 or below, and not on active duty in the military, whose responses to CI items provided estimates representative of all civilian U.S. adults.

In addition to the major topical components, the NHES:96 Screener collected demographic and educational information on all members in every household contacted, whether or not anyone in the household was selected for an extended interview. (The term "extended interview" refers to the interview pertaining to the topical component of the study, that is, the Parent PFI/CI, the Youth CI, or the Adult CI interview.)

This manual, the *NHES:96 Data File User's Manual, Volume V: Adult Civic Involvement*, provides documentation and guidance for users of the public release data file for the Adult Civic Involvement (Adult CI) component of the 1996 National Household Education Survey (NHES:96). This volume contains a description of the Adult CI data file and a discussion of data considerations and anomalies. Included as appendixes are the public file layout, SAS code for creating derived variables, the codebook for the Adult CI public data file, and directions and sample code for linking NHES:96 data files.

Volume V is meant to be read in conjunction with the *NHES:96 Data File User's Manual, Volume I*. More information about the purpose of the study, the sample design, the other survey components, the data collection instruments, and data collection and data processing procedures is contained in the *NHES:96 Data File User's Manual, Volume I*. Information about the Household & Library public data file can be found in Volume II, and information about the Parent and Family

Involvement in Education and Civic Involvement (Parent PFI/CI) public data file and the Youth Civic Involvement (Youth CI) public data file can be found in Volumes III and IV of the manual, respectively.

6. GUIDE TO THE DATA FILE AND CODEBOOK

6.1 Content and Organization of the Data File

This section describes the content of the Adult CI public release data file of the NHES:96. This file contains data from all completed Adult CI interviews. There are two records for each Adult CI interview completed, so the file contains 4,500 records for the 2,250 cases. The file is organized so that logically related sets of variables are grouped together. The data items are listed in the file in the following order: system variables, household membership variables, questionnaire item variables, household characteristics variables, derived variables, weighting and variance estimation variables, and imputation flag variables.

A list of all the variables in the Adult CI data file is shown in appendix B. The VARIABLE NAME column displays the unique identifier for each variable in the data file. The VARIABLE LABEL column displays a short description associated with the variable as it appears in the SAS system file. The FORMAT column indicates if a variable has a numeric ("N") or a character ("A") format. All of the variables in the Adult CI data file have numeric formats, with the exception of SGRADE. The RECORD NUMBER column indicates whether the variable is located on the first or second record. The LENGTH column indicates the length of the variable by the number of digits. The length descriptor also includes the number of digits found after the decimal point for noninteger numeric variables (e.g., weight variables). The position of the variable is indicated in the START and END columns which indicate the position in the file where the variable begins and ends.

A value of "-1" for any variable in the data file indicates that a case was part of a legitimate skip. For example, if the adult learned English as his/her first language (ALANG, R17), the question about what language he/she speaks most at home (ASPEAK, R18) would have a value of -1, because the question was inapplicable. This convention of assigning -1 to all legitimate skips applies to all NHES data files.

The NHES:96 data files are provided on CD-ROM and are accessible through an Electronic CodeBook (ECB) that allows data users to view variable frequencies, tag variables for extraction, and create the SAS, SPSS for DOS, or SPSS for Windows code needed to create an extract file for analysis purposes. The ECB contains all of the NHES:96 data sets: the Household & Library file, the Parent PFI/CI file, the Youth CI file, and the Adult CI file. The ECB also contains all data sets for previous NHES collections and documentation for every file. Instructions for using the CD-ROM and ECB are provided in a separate document, the *National Household Education Survey: NHES:91/93/95/96 Electronic CodeBook (ECB) User's Guide* (Collins and Chandler forthcoming). The sections that follow describe the contents of the Adult CI data file.

6.1.1 System Variables

System variables are created during the conduct of an interview and are instrumental in the successful administration of the interview. Their creation is transparent to the interviewer and to the respondent. System variables fall into two categories: linking variables (record identifiers or ID numbers) and interview status variables. Linking variables are record identifiers that provide a link to

other interviews completed in the same the same household. Status variables are set at the completion of each interview to define interview status.

BASMID is the unique 12-digit identifier variable for the interview. It is composed of the eight-digit household identifier, the interview subject's two-digit household-member person number, and the two-digit number, 03, that indicates that the interview was an Adult CI interview.

BASEID is the eight-digit identifier for the household. This ID number also forms the first eight digits of interview ID numbers for other interviews in the household, providing a means of linking interviews within the same household. See appendix E for instructions for linking the NHES:96 data files.

ENGLSPAN is the variable that indicates whether the interview was conducted in English or in Spanish.

The values for ENGLSPAN are:

- 1 = Interview was conducted in English
- 2 = Interview was conducted in Spanish

6.1.2 Household Membership Variables

All household members were enumerated in the Screener. Data collected included age and sex (S6), educational status (SX7 through SX14), and demographic characteristics (SX15 through SX22 and SX27 through SX33OV). The gender collected during the household enumeration in the Screener (S6) was used to drive the gender-based wording of subsequent questions throughout the extended interview.

PERSNUM is the household member person number of the respondent.

AGE is the respondent's age.

SEX is the respondent's sex.

RACE indicates the respondent's race.

HISPANIC indicates whether the respondent is Hispanic.

OTHRAC indicates the respondent's race if "Some other race" was reported at RACE (SX21).

SENROLL indicates whether the respondent is attending or enrolled in school.

SGRADE indicates the grade or year of school the respondent is attending.

MARITL indicates the marital status of the respondent.

BORNUS indicates whether the respondent was born in one of the 50 States or the District of Columbia, a U.S. territory, or some other country.

6.1.3 Questionnaire Item Variables

The questionnaire item variables appear on the file in the same order as they were asked. Refer to the questionnaire in Volume I, appendix A for the order. The items on language and highest grade completed appear in the Screener questionnaire and the Adult CI questionnaire. The Adult CI responses have been retained, since they are responses given by the respondent. Some variables were excluded from the file for confidentiality reasons. These include the names of household members, verbatim string responses that might identify persons or places, and the individual ZIP Codes. Some of these variables are included in a separate restricted-use data tile (see section 6.3 below). The Adult CI questionnaire appears with the Screener, the Parent PFI/CI, and the Youth CI questionnaires, in Volume I, appendix A; variable names are provided to the left of each question. Those followed by "/R" appear only on a restricted-use data file that may be obtained through a special licensing agreement with NCES.

If a value for a response option is found in the questionnaire, but not found the frequency, no respondent selected that response. The only variable that meet this condition in the Adult CI data file is AGRAD1 (the specific highest grade attended for those whose highest grade is grade 8 or less).

6.1.4 Household Characteristics Variables

Household characteristics variables are variables that reflect characteristics of the household as a unit. For example, questions were asked about whether the home was owned or rented and the type and size of community where the household was located. These household items appear on the file in the same order as they were asked. Some of the household questions were asked at the end of the first extended interview in the household. Because they are actually part of the Screener, refer to the questionnaire in Volume I, appendix A for the questions and their order.

6.1.5 Derived Variables

Derived variables were developed and included in the Adult CI public use data file to aid users in their analyses. The derived variables fall into two categories: questionnaire item variables and counter variables. Questionnaire item-derived variables were created by combining two or more items from the questionnaire. Counter-derived variables were created by counting the number of persons with specific characteristics enumerated in the household. Linked-derived variables, created by using the respondent's ZIP Code to extract data from the 1990 Census of Population Summary Tape File 3B (STF3B), are available on a restricted-use file.

The derived variables appear together on the file in their own section in alphabetical order. They are listed below in the same order with an explanation of how they were derived. The actual SAS code to create these variables is found in appendix C, with the exception of counter variables and CENREG (Census region).

CENREG is a derived variable that identifies the Census region in which the subject adult lives. This variable is created by grouping states.

The following states and the District of Columbia are in each Census region:

Northeast: CT, MA, ME, NH, NJ, NY, PA, RI, VT
South: AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV
Midwest: IA, IL, IN, KS, MI, MN, MO, ND, NE, OH, SD, WI
West: AK, AZ, CA, CO, HI, ID, MT, NM, NV, OR, UT, WA, WY

The values for CENREG are:

1 = Northeast
2 = South
3 = Midwest
4 = West

COMMUNITY is a derived variable representing the respondent's report of the size of the community in which the household is located. **COMMUNITY** was created from the variables HCCOMMUN (SX31), HCSUB (SX31OV), and HCCITY (SX31OV2).

The values for **COMMUNITY** are:

1 = Very large city (over 500,000 people)
2 = Large city (100,000 to 500,000 people)
3 = Medium sized city (50,000 to 100,000 people)
4 = Suburb of a very large city
5 = Suburb of a large city
6 = Suburb of a medium city
7 = Small city or town of fewer than 50,000 people that is not a suburb of a larger city
8 = Rural or farming community

EMPLOYED is a derived variable that identifies the work status of the sampled adult. **EMPLOYED** was created from the variables AWORK (R21), ALEAVE (R22), AHOURS (R23), and APUBL, APRIV, AEMPL, AREL, and AANSAD (all from R25).

The values for EMPLOYED are:

- 1 = Working 35 hours or more per week
- 2 = Working less than 35 hours per week
- 3 = Looking for work
- 4 = Not in the labor force

HH18OVER is the counter-derived variable that indicates the number of household members age 18 and older. The screener responses to HHAGE1 through HHAGE16 (S6), found in the Household & Library file, were counted for this variable.

HHTOTAL is the counter-derived variable that indicates the total number of household members. The screener responses to HHAGE1 through HHAGE16 (S6), found in the Household & Library file, were examined for this variable.

HHUNDR18 is the counter-derived variable that indicates the number of household members younger than 18 years old. The screener responses to HHAGE1 through HHAGE16 (S6), found in the Household & Library file, were examined for this variable.

HIGHGRAD indicates the highest level of education of any household member. HIGHGRAD was created using the variables SGRADE1 through SGRADE16 (SX9), GRADE1 through GRADE16 (SX13), and SDIPL1 through SDIPL16 (SX14)(S6), found in the Household & Library file.

The values for HIGHGRAD are:

- 1 = Less than a high school diploma
- 2 = High school diploma or equivalent
- 3 = Vocational or technical degree or some college
- 4 = Bachelor's degree
- 5 = Graduate or professional school

RACEETHN denotes both the race and ethnicity of the adult. If the respondent designates his/her ethnicity as Hispanic, RACEETHN is Hispanic regardless of whether RACE was classified as white, black, or another race. RACEETHN was created using the variables RACE (SX21), OTHRAC (SX21A), and HISPANIC (SX22).

The values for RACEETHN are:

- 1 = White, non-Hispanic
- 2 = Black, non-Hispanic
- 3 = Hispanic
- 4 = All other races (e.g., American Indian or Alaska Native, Asian or Pacific Islander), non-Hispanic

6.1.6 Weighting and Variance Estimation Variables

The first variable in this section of the file is FAWT. It is the variable that should be used as the weight variable to estimate the characteristics of adults. This weight contains all of the adjustments for the probabilities of selection, nonresponse, and undercoverage as described in Volume I, chapter 3.

The 80 replicate weights, FAWTR1 to FAWTR80, are the next variables in this section. These replicate weights can be used with the WesVarPC Windows-based software program to produce estimates of the sampling errors of the estimates. More details on how the replicate weights were created and how they can be used with WesVarPC are given in Volume I, chapter 3, along with an approximation method that does not involve using the WesVarPC procedure.

The remaining two variables in this section are ASTRATUM and APSU. These variables are provided to enable users to compute sampling errors using Taylor Series approximations, such as the SUDAAN procedure. The methods used to construct the values for ASTRATUM and APSU are also discussed in Volume I, chapter 3.

6.1.7 Imputation Flag Variables

Item nonresponse occurred when some, but not all, of the responses were missing from an otherwise cooperating respondent. For all the items on the Adult CI public use file, except the government knowledge questions (R14a-e and R15a-e), the missing data were imputed, or "filled in," to help users of the data. For each variable involved in imputation, an imputation flag variable was created. If the imputation flag is equal to 0, then no imputation was performed on that case. This flag can be used to identify imputed values. Volume I, section 3.8 discusses the meaning of values assigned to the imputation flags.

The naming convention for the imputation flag variables was to drop the last letter of the variable name and replace it with an "f." For example, the imputation flag for SEX is SEF. The imputation flags appear on the file in the same order as the items appear in the questionnaire. This naming convention holds true for all Adult CI variables except for variables that will become confused with other variables when the last letter is dropped or variables that end in a number. In these cases, the letter before the last digit is dropped and replaced with an "f." For example, the imputation flag for CAVOTE5 (R10) is CAVOTF5. The imputation flags appear on the file in the same order as the items.

6.1.8 Numeric and Character Variables

All of the variables in the Adult CI file have numeric formats, except SGRADE (SX9).

6.2 Guide to the Codebook

The codebook, shown in appendix D, contains complete descriptions of the contents of the data file. The codebook contains system variables, household membership variables, questionnaire variables, household characteristics variables, derived variables, weighting and variance estimation

variables, and imputation flag variables. The codebook provides all the pertinent information for the variables in the file, including the variable name, the question wording, the position and format of the variable in the file, and the responses to the item. The unweighted frequency, unweighted percent, and weighted percent are provided along with each response. Figure 6-1 provides a description of each of the items appearing in the codebook.

6.3 Public and Proprietary Data Files

This manual is designed to assist users of the public use the Adult CI data file. The public use file contains all the variables detailed above but does not contain certain variables excluded from the file for confidentiality reasons. These include the names of household members, verbatim string responses that might identify persons, and respondents' individual ZIP Codes. Some of these variables (e.g., verbatim strings of other-specify categories) that are excluded from the public file are included on a separate proprietary, or restricted-use, file. These variables are indicated with a "/R" on the Adult CI questionnaire in Volume I, appendix A. The Adult CI proprietary data file also contains close to 100 ZIP code variables from the 1990 Census of Population Summary Tape File 3B (STF3B), including the median household income of the area, the level of community mobility in the area, and the percentage of owner-occupied households in the area. The proprietary data files may be obtained through a special licensing agreement with NCES. Contact NCES for details on how to become licensed.

6.4 Linking the Household & Library File to Other NHES:96 Data Files

It is possible to link the Household & Library file to the Parent PFI/CI, the Youth CI, and the Adult CI data files. Instructions for doing so are located in appendix E.

Figure 6-1.—Example of the codebook format

- (1) CAPARENT = (2) R16-PARENT OF CHILD 18 OR UNDER
 (3) R16 Are you a parent, stepparent, or guardian of (this child/any of these children)?
 (4) RECORD: 1 POSITION: 103-104
 (5) FORMAT: N2

(6) RESPONSE	(7) CODES	(8) FREQ	(9) UNWGTD PERCENT	(10) WGTD PERCENT
1 YES	1	863	38.4%	91.1%
2 NO	2	54	2.4%	8.9%
RESERVED CODES				
-1 INAPPLICABLE	-1	1333	59.2%	(MISS)
TOTALS		2250	100.0%	100.0%

DESCRIPTIONS:

- (1) Variable name: This is the variable name associated with each item. This is the unique identifier present in the SAS or SPSS data file.
- (2) Variable label: A short label, which is associated with each of the variables, is presented here. This label appears in the SAS or SPSS data file. Labels contain the questionnaire item numbers. Labels that begin with the letter "D" indicate a derived variable.
- (3) Question wording: This is the exact question wording as it appeared in the questionnaire.
- (4) Record and position: These provide the record number (1 or 2) and the starting and ending position of the variable in the raw data file on tape.
- (5) Format: This provides the variable type, its width, and the number of positions after the decimal point, if necessary. A data type of "N" represents numeric variables and "A" represents character variables. In this example, CAPARENT is a numeric variable with a length of 2.
- (6) Response categories: This column provides the response categories for the variable.
- (7) Response codes: This column provides the actual numeric/alphanumeric codes present in the data files.
- (8) Unweighted frequency counts: This column displays the unweighted frequency counts for this variable. The counts for missing values will also be included for the unweighted values, but not for the weighted values.
- (9) Unweighted percentages: This column displays the unweighted frequency counts from the previous column as percentages. This column will also contain percentages for missing values.
- (10) Weighted percentages: This column displays the percentages of frequency counts weighted up to the population. This column will not include percentages for missing values.

7. DATA CONSIDERATIONS AND ANOMALIES

The purpose of this section is to bring to the user's attention certain data considerations and data anomalies in the NHES:96 Adult CI data; to describe the nature of those anomalies; and, where appropriate, to identify possible means of taking them into account when analyzing the Adult CI data. Only two data considerations that users should be aware of have been identified in the Adult CI data file.

7.1 Type of Community in Which the Household is Located

The NHES:96 Adult CI file includes a measure of the urbanicity of the sampled household, COMMUNITY. This variable was copied from the Household & Library data file to the corresponding record on the Adult CI file. The creation of the derived variable COMMUNITY was described in chapter 6. At the household level, the item response rates for the variables used to create COMMUNITY were somewhat lower than most variables in the NHES. HCCOMMUN (SX31, the type of community) had an item response rate of 92.5 percent; HCSUB (SX31OV, size of city to which a suburb belonged, if applicable) had an item response rate of 93.1 percent; and HCCITY (SX31OV2, size of the city, if applicable) had an item response rate of 95.7 percent. When these three variables were combined to create the derived variable COMMUNITY, 85.3 percent had unimputed responses for all three variables. Some respondents simply do not know the size of the community in which they live. This suggests the possibility that some who responded to the questions may have guessed, although this cannot be measured directly. Analysts should keep this in mind when using the variable COMMUNITY, as this variable may contain response error. The component variables of COMMUNITY are also included in the Adult CI data file.

7.2 Income to the Nearest Thousand Dollars

In those households whose income category and household size indicated that they may be at or below the poverty line, household income to the nearest thousand dollars was requested. As the values in the data file show, some respondents did not answer in thousands, but gave somewhat more specific answers. Rather than lose this information, the exact response was retained.

This page is blank.

APPENDIX B

ADULT CI PUBLIC FILE LAYOUT IN POSITION ORDER

This page is blank.

Adult CI Public File Layout in Position Order

VARIABLE NAME	VARIABLE LABEL	FORMAT	RECORD NUMBER	LENGTH	START COLUMN	END COLUMN
BASPID	INTERVIEW ID NUMBER	N	1	12	1	12
BASEID	HOUSEHOLD ID NUMBER	N	1	8	13	20
ENGLSPAN	EXTENDED IN ENGLISH OR SPANISH	N	1	2	21	22
PERSNUM	PERSON'S ENUMERATION NUMBER	N	1	2	23	24
AGE	S6-AGE	N	1	2	25	26
SEX	S6-SEX	N	1	2	27	28
RACE	SX21-RACE	N	1	2	29	30
HISPANIC	SX22-HISPANIC	N	1	2	31	32
OTHRAC	SX21A-OTHER RACE CATEGORY	N	1	2	33	34
SENROLL	SX7-ATTENDING/ENROLLED IN SCH	N	1	2	35	36
SGRADE	SX9-GRADE/YEAR OF SCH ATTENDING	A	1	2	37	38
MARITL	SX15-MARITAL STATUS	N	1	2	39	40
BORNUS	SX19-WHAT COUNTRY BORN	N	1	2	41	42
CARDPAPR	R1-FREQ READ NEWSPAPER	N	1	2	43	44
CARDMAGS	R2-# DIFFERENT MAGS LOOK AT OR READ REG	N	1	2	45	46
CARDBOOK	R3-READ ANY BOOKS PAST 6 MO	N	1	2	47	48
CARDNEUW	R4-FREQ ADULT READS NATL NEWS	N	1	2	49	50
CAWATCHU	R5-FREQ ADULT WATCH/LSTN NATL NEWS	N	1	2	51	52
CAOTHORG	R6-BELONGS TO ANY ORGNZTN	N	1	2	53	54
CARELFRQ	R7-FREQ ATTNDDED REL SERV PST YR	N	1	2	55	56
CASERV	R8-DOES COMMUNITY SERV	N	1	2	57	58
CAMONEY	R9A-GAVE \$ TO POLITICAL CAUSE	N	1	2	59	60
CAVOLUNT	R9B-WORKED FOR POLITICAL CAUSE	N	1	2	61	62
CATELISS	R9C-CONTACTED OFCL ABT ISSUE	N	1	2	63	64
CAPUBMTG	R9D-ATTNDDED PUBLIC MTG	N	1	2	65	66
CABOYCOT	R9E-PARTICIPATED IN PROTEST/BOYCT	N	1	2	67	68
CAVOTE5	R10-VOTED IN LAST 5 YEARS	N	1	2	69	70
CACOMPLI	R11A-CAN'T UNDERSTAND POLITICS/GOVT	N	1	2	71	72
CAFAMSAY	R11B-FAM HAS NO SAY IN WHAT GOVT DOES	N	1	2	73	74
CAAGNST	R11C-ALLOW FREEDOM TO SPEAK AGNST RELGN	N	1	2	75	76
CABOOK	R11D-SOME BOOKS SHLD BE KPT OUT/PUB LIB	N	1	2	77	78
CALETTER	R12-COULD WRITE LETTER TO GOVT OFCL	N	1	2	79	80
CAMTG	R13-COULD MAKE STATEMENT AT PUBLIC MTG	N	1	2	81	82
CAVP	R14A-JOB/POL OFFICE HELD BY AL GORE	N	1	2	83	84
CALAW	R14B-WHO DETERMINES LAW CONSTITUTIONAL	N	1	2	85	86
CAHOUSE	R14C-PARTY W/MOST MEMBRS IN HOUSE	N	1	2	87	88
CAVETO	R14D-MAJORITY NEEDED TO OVERRIDE VETO	N	1	2	89	90
CACONSERV	R14E-PARTY MORE CONSERV/NATL LEVEL	N	1	2	91	92
CASPKR	R15A-JOB/POL OFC HELD BY NEWT GINGRICH	N	1	2	93	94
CAJUDGE	R15B-WHO NOMINATES FED COURT JUDGES	N	1	2	95	96
CASENATE	R15C-PARTY W/MOST MEMBRS IN SENATE	N	1	2	97	98
CACONST	R15D-1ST 10 AMENDMENTS TO CONSTIT	N	1	2	99	100
CADFENS	R15E-PARTY FAVORS LRGR DEFENSE BUDGET	N	1	2	101	102
CAPARENT	R16-PARENT OF CHLD IN HH 18 OR YOUNGER	N	1	2	103	104
ALANG	R17-1ST LANG SPOKEN	N	1	2	105	106
ASPEAK	R18-LANG SPOKEN MOST AT HOME	N	1	2	107	108
AGRADE	R19-HIGHEST GRADE COMPLETED	N	1	2	109	110
AGRAD1	R19-ACTUAL GRADE 0-8 COMPLETED	N	1	2	111	112
AGRAD2	R19-ACTUAL GRADE 9-11 COMPLETED	N	1	2	113	114
ADIPL	R20-HAS HS DIPLOMA/GED	N	1	2	115	116
AWORK	R21-WORKED FOR PAY LAST WEEK	N	1	2	117	118
ALEAVE	R22-ON LEAVE OR VACATION LAST WEEK	N	1	2	119	120
AHOURS	R23-HRS/WK USUALLY WORK FOR PAY	N	1	2	121	122
AMTHS	R24-MONTHS WORKED IN PAST YR	N	1	2	123	124
ALOOK	R25-LOOKING FOR WORK PAST 4 WKS	N	1	2	125	126
APUBL	R26-CHECKED W/PUBLIC EMPLOY AGENCY	N	1	2	127	128
APRIV	R26-CHECKED W/PRIVATE EMPLOY AGENCY	N	1	2	129	130
AEMPL	R26-CHECKED W/EMPLOYER DIRECTLY	N	1	2	131	132
AREL	R26-CHECKED W/FRIENDS/RELATIVES	N	1	2	133	134
AANSAD	R26-PLACED/ANSWERED ADS	N	1	2	135	136
AREAD	R26-READ WANT ADS	N	1	2	137	138
AOTHER	R26-DID OTHR THINGS TO FIND WORK	N	1	2	139	140
AACTY	R27-MAIN ACTIVITY LAST WEEK	N	1	2	141	142
ADISCIP	R28A-STRICTR DISCIPLINE WLD IMPROVE EDUC	N	1	2	143	144
ASTANDS	R28B-PROMOTION STANDARDS WLD IMPROVE EDU	N	1	2	145	146
AEVAL	R28C-TCHR EVALUATIONS WLD IMPROVE EDUC	N	1	2	147	148
ASCHLYR	R28D-LONGER SCH YR WLD IMPROVE EDUC	N	1	2	149	150

VARIABLE NAME	VARIABLE LABEL	FORMAT	RECORD NUMBER	LENGTH	START COLUMN	END COLUMN
XHHBORN	SX17-ALL IN HH BORN IN US	N	1	2	151	152
XHHLANG	SX18-ALL IN HH LEARN ENGL/1ST LANG	N	1	2	153	154
HOWNHOME	SX27-OWN, RENT HOME/OTHR ARRNGMNT	N	1	2	155	156
HCCOMMUN	SX31-COMMUNITY DESCRIPTION	N	1	2	157	158
HCSUB	SX31OV-SIZE OF SUBURB	N	1	2	159	160
HCCITY	SX31OV2-SIZE OF CITY	N	1	2	161	162
HINCMRNG	SX33- TOTAL HH INCOME RANGE	N	1	2	163	164
HINCOME	SX33-TOTAL HH INCOME RANGE 2	N	1	2	165	166
HINCMEXT	SX33OV-EXACT HH INC NEAREST \$1000	N	1	5	167	171
CENREG	D-CENSUS REGION	N	1	2	172	173
COMMUNITY	D-SIZE OF COMMUNITY	N	1	2	174	175
EMPLOYED	D-WORK STATUS	N	1	2	176	177
HH18OVER	D-NUMBER OF HH MMBRS AGE 18 AND OLDER	N	1	2	178	179
HHTOTAL	D-TOTAL NUMBER OF HH MEMBERS	N	1	2	180	181
HHUNDR18	D-NUMBER OF HH MMBRS AGE 17 AND YOUNGER	N	1	2	182	183
HIGHGRAD	D-HIGHEST EDUCATION OF ANY HH MEMBER	N	1	2	184	185
RACEETHN	D-RACE-ETHNICITY	N	1	2	186	187
FAWT	FINAL (RAKED) ADULT INTERVIEW WGT	N	1	10.3	188	197
BASMID	INTERVIEW ID NUMBER	N	2	12	1	12
FAWTR1	FINAL (RAKED) ADULT INTV. REPL. WGT-1	N	2	10.3	13	22
FAWTR2	FINAL (RAKED) ADULT INTV. REPL. WGT-2	N	2	10.3	23	32
FAWTR3	FINAL (RAKED) ADULT INTV. REPL. WGT-3	N	2	10.3	33	42
FAWTR4	FINAL (RAKED) ADULT INTV. REPL. WGT-4	N	2	10.3	43	52
FAWTR5	FINAL (RAKED) ADULT INTV. REPL. WGT-5	N	2	10.3	53	62
FAWTR6	FINAL (RAKED) ADULT INTV. REPL. WGT-6	N	2	10.3	63	72
FAWTR7	FINAL (RAKED) ADULT INTV. REPL. WGT-7	N	2	10.3	73	82
FAWTR8	FINAL (RAKED) ADULT INTV. REPL. WGT-8	N	2	10.3	83	92
FAWTR9	FINAL (RAKED) ADULT INTV. REPL. WGT-9	N	2	10.3	93	102
FAWTR10	FINAL (RAKED) ADULT INTV. REPL. WGT-10	N	2	10.3	103	112
FAWTR11	FINAL (RAKED) ADULT INTV. REPL. WGT-11	N	2	10.3	113	122
FAWTR12	FINAL (RAKED) ADULT INTV. REPL. WGT-12	N	2	10.3	123	132
FAWTR13	FINAL (RAKED) ADULT INTV. REPL. WGT-13	N	2	10.3	133	142
FAWTR14	FINAL (RAKED) ADULT INTV. REPL. WGT-14	N	2	10.3	143	152
FAWTR15	FINAL (RAKED) ADULT INTV. REPL. WGT-15	N	2	10.3	153	162
FAWTR16	FINAL (RAKED) ADULT INTV. REPL. WGT-16	N	2	10.3	163	172
FAWTR17	FINAL (RAKED) ADULT INTV. REPL. WGT-17	N	2	10.3	173	182
FAWTR18	FINAL (RAKED) ADULT INTV. REPL. WGT-18	N	2	10.3	183	192
FAWTR19	FINAL (RAKED) ADULT INTV. REPL. WGT-19	N	2	10.3	193	202
FAWTR20	FINAL (RAKED) ADULT INTV. REPL. WGT-20	N	2	10.3	203	212
FAWTR21	FINAL (RAKED) ADULT INTV. REPL. WGT-21	N	2	10.3	213	222
FAWTR22	FINAL (RAKED) ADULT INTV. REPL. WGT-22	N	2	10.3	223	232
FAWTR23	FINAL (RAKED) ADULT INTV. REPL. WGT-23	N	2	10.3	233	242
FAWTR24	FINAL (RAKED) ADULT INTV. REPL. WGT-24	N	2	10.3	243	252
FAWTR25	FINAL (RAKED) ADULT INTV. REPL. WGT-25	N	2	10.3	253	262
FAWTR26	FINAL (RAKED) ADULT INTV. REPL. WGT-26	N	2	10.3	263	272
FAWTR27	FINAL (RAKED) ADULT INTV. REPL. WGT-27	N	2	10.3	273	282
FAWTR28	FINAL (RAKED) ADULT INTV. REPL. WGT-28	N	2	10.3	283	292
FAWTR29	FINAL (RAKED) ADULT INTV. REPL. WGT-29	N	2	10.3	293	302
FAWTR30	FINAL (RAKED) ADULT INTV. REPL. WGT-30	N	2	10.3	303	312
FAWTR31	FINAL (RAKED) ADULT INTV. REPL. WGT-31	N	2	10.3	313	322
FAWTR32	FINAL (RAKED) ADULT INTV. REPL. WGT-32	N	2	10.3	323	332
FAWTR33	FINAL (RAKED) ADULT INTV. REPL. WGT-33	N	2	10.3	333	342
FAWTR34	FINAL (RAKED) ADULT INTV. REPL. WGT-34	N	2	10.3	343	352
FAWTR35	FINAL (RAKED) ADULT INTV. REPL. WGT-35	N	2	10.3	353	362
FAWTR36	FINAL (RAKED) ADULT INTV. REPL. WGT-36	N	2	10.3	363	372
FAWTR37	FINAL (RAKED) ADULT INTV. REPL. WGT-37	N	2	10.3	373	382
FAWTR38	FINAL (RAKED) ADULT INTV. REPL. WGT-38	N	2	10.3	383	392
FAWTR39	FINAL (RAKED) ADULT INTV. REPL. WGT-39	N	2	10.3	393	402
FAWTR40	FINAL (RAKED) ADULT INTV. REPL. WGT-40	N	2	10.3	403	412
FAWTR41	FINAL (RAKED) ADULT INTV. REPL. WGT-41	N	2	10.3	413	422
FAWTR42	FINAL (RAKED) ADULT INTV. REPL. WGT-42	N	2	10.3	423	432
FAWTR43	FINAL (RAKED) ADULT INTV. REPL. WGT-43	N	2	10.3	433	442
FAWTR44	FINAL (RAKED) ADULT INTV. REPL. WGT-44	N	2	10.3	443	452
FAWTR45	FINAL (RAKED) ADULT INTV. REPL. WGT-45	N	2	10.3	453	462
FAWTR46	FINAL (RAKED) ADULT INTV. REPL. WGT-46	N	2	10.3	463	472
FAWTR47	FINAL (RAKED) ADULT INTV. REPL. WGT-47	N	2	10.3	473	482
FAWTR48	FINAL (RAKED) ADULT INTV. REPL. WGT-48	N	2	10.3	483	492
FAWTR49	FINAL (RAKED) ADULT INTV. REPL. WGT-49	N	2	10.3	493	502
FAWTR50	FINAL (RAKED) ADULT INTV. REPL. WGT-50	N	2	10.3	503	512
FAWTR51	FINAL (RAKED) ADULT INTV. REPL. WGT-51	N	2	10.3	513	522

VARIABLE NAME	VARIABLE LABEL	FORMAT	RECORD NUMBER	LENGTH	START COLUMN	END COLUMN
FAWTR52	FINAL (RAKED) ADULT INTV. REPL. WGT-52	N	2	10.3	523	532
FAWTR53	FINAL (RAKED) ADULT INTV. REPL. WGT-53	N	2	10.3	533	542
FAWTR54	FINAL (RAKED) ADULT INTV. REPL. WGT-54	N	2	10.3	543	552
FAWTR55	FINAL (RAKED) ADULT INTV. REPL. WGT-55	N	2	10.3	553	562
FAWTR56	FINAL (RAKED) ADULT INTV. REPL. WGT-56	N	2	10.3	563	572
FAWTR57	FINAL (RAKED) ADULT INTV. REPL. WGT-57	N	2	10.3	573	582
FAWTR58	FINAL (RAKED) ADULT INTV. REPL. WGT-58	N	2	10.3	583	592
FAWTR59	FINAL (RAKED) ADULT INTV. REPL. WGT-59	N	2	10.3	593	602
FAWTR60	FINAL (RAKED) ADULT INTV. REPL. WGT-60	N	2	10.3	603	612
FAWTR61	FINAL (RAKED) ADULT INTV. REPL. WGT-61	N	2	10.3	613	622
FAWTR62	FINAL (RAKED) ADULT INTV. REPL. WGT-62	N	2	10.3	623	632
FAWTR63	FINAL (RAKED) ADULT INTV. REPL. WGT-63	N	2	10.3	633	642
FAWTR64	FINAL (RAKED) ADULT INTV. REPL. WGT-64	N	2	10.3	643	652
FAWTR65	FINAL (RAKED) ADULT INTV. REPL. WGT-65	N	2	10.3	653	662
FAWTR66	FINAL (RAKED) ADULT INTV. REPL. WGT-66	N	2	10.3	663	672
FAWTR67	FINAL (RAKED) ADULT INTV. REPL. WGT-67	N	2	10.3	673	682
FAWTR68	FINAL (RAKED) ADULT INTV. REPL. WGT-68	N	2	10.3	683	692
FAWTR69	FINAL (RAKED) ADULT INTV. REPL. WGT-69	N	2	10.3	693	702
FAWTR70	FINAL (RAKED) ADULT INTV. REPL. WGT-70	N	2	10.3	703	712
FAWTR71	FINAL (RAKED) ADULT INTV. REPL. WGT-71	N	2	10.3	713	722
FAWTR72	FINAL (RAKED) ADULT INTV. REPL. WGT-72	N	2	10.3	723	732
FAWTR73	FINAL (RAKED) ADULT INTV. REPL. WGT-73	N	2	10.3	733	742
FAWTR74	FINAL (RAKED) ADULT INTV. REPL. WGT-74	N	2	10.3	743	752
FAWTR75	FINAL (RAKED) ADULT INTV. REPL. WGT-75	N	2	10.3	753	762
FAWTR76	FINAL (RAKED) ADULT INTV. REPL. WGT-76	N	2	10.3	763	772
FAWTR77	FINAL (RAKED) ADULT INTV. REPL. WGT-77	N	2	10.3	773	782
FAWTR78	FINAL (RAKED) ADULT INTV. REPL. WGT-78	N	2	10.3	783	792
FAWTR79	FINAL (RAKED) ADULT INTV. REPL. WGT-79	N	2	10.3	793	802
FAWTR80	FINAL (RAKED) ADULT INTV. REPL. WGT-80	N	2	10.3	803	812
APSU	FOR USE IN TAYLOR SERIES VARIANCE	N	2	4	813	816
ASTRATUM	FOR USE IN TAYLOR SERIES VARIANCE	N	2	2	817	818
AGF	IMPUTATION FLAG	N	2	2	819	820
SEF	IMPUTATION FLAG	N	2	2	821	822
RACF	IMPUTATION FLAG	N	2	2	823	824
HISPANIF	IMPUTATION FLAG	N	2	2	825	826
OTHRAF	IMPUTATION FLAG	N	2	2	827	828
SENROLF	IMPUTATION FLAG	N	2	2	829	830
SGRADF	IMPUTATION FLAG	N	2	2	831	832
MARITF	IMPUTATION FLAG	N	2	2	833	834
BORNUF	IMPUTATION FLAG	N	2	2	835	836
CARDPAPF	IMPUTATION FLAG	N	2	2	837	838
CARDMAGF	IMPUTATION FLAG	N	2	2	839	840
CARDNEWF	IMPUTATION FLAG	N	2	2	841	842
CAWATCHF	IMPUTATION FLAG	N	2	2	843	844
CAOTHORF	IMPUTATION FLAG	N	2	2	845	846
CARELFRF	IMPUTATION FLAG	N	2	2	847	848
CASERVF	IMPUTATION FLAG	N	2	2	849	850
CAMONEF	IMPUTATION FLAG	N	2	2	851	852
CAVOLUNF	IMPUTATION FLAG	N	2	2	853	854
CATELISF	IMPUTATION FLAG	N	2	2	855	856
CAPUBMTF	IMPUTATION FLAG	N	2	2	857	858
CABOYCOF	IMPUTATION FLAG	N	2	2	859	860
CAVOTF5	IMPUTATION FLAG	N	2	2	861	862
CACOMPLF	IMPUTATION FLAG	N	2	2	863	864
CAFAMSAF	IMPUTATION FLAG	N	2	2	865	866
CAAGNSF	IMPUTATION FLAG	N	2	2	867	868
CABOOF	IMPUTATION FLAG	N	2	2	869	870
CALETTTF	IMPUTATION FLAG	N	2	2	871	872
CAMTF	IMPUTATION FLAG	N	2	2	873	874
CAPARENF	IMPUTATION FLAG	N	2	2	875	876
ALANF	IMPUTATION FLAG	N	2	2	877	878
ASPEAF	IMPUTATION FLAG	N	2	2	879	880
AGRADF	IMPUTATION FLAG	N	2	2	881	882
AGRAF1	IMPUTATION FLAG	N	2	2	883	884
ADIPF	IMPUTATION FLAG	N	2	2	885	886
AWORF	IMPUTATION FLAG	N	2	2	887	888
AHOURF	IMPUTATION FLAG	N	2	2	889	890
AMTHF	IMPUTATION FLAG	N	2	2	891	892
AACTF	IMPUTATION FLAG	N	2	2	893	894
ADISCIF	IMPUTATION FLAG	N	2	2	895	896

VARIABLE NAME	VARIABLE LABEL	FORMAT	RECORD NUMBER	LENGTH	START COLUMN	END COLUMN
ASTANDF	IMPUTATION FLAG	N	2	2	897	898
AEVAF	IMPUTATION FLAG	N	2	2	899	900
ASCHLYF	IMPUTATION FLAG	N	2	2	901	902
XHHBORF	IMPUTATION FLAG	N	2	2	903	904
XHHLANF	IMPUTATION FLAG	N	2	2	905	906
HOWNHOMF	IMPUTATION FLAG	N	2	2	907	908
HINCMRNF	IMPUTATION FLAG	N	2	2	909	910
HINCOMF	IMPUTATION FLAG	N	2	2	911	912
HINCMEXF	IMPUTATION FLAG	N	2	2	913	914
HCCOMMUF	IMPUTATION FLAG	N	2	2	915	916
HCSUF	IMPUTATION FLAG	N	2	2	917	918
HCCITF	IMPUTATION FLAG	N	2	2	919	920

NOTE: The variables RECNUM is located in the last column of each record (column 1,024). The value of RECNUM varies with the record number of a given case. RECNUM is set to one on the first record of every case, 2 for the second record, and 3 for the third record. Each case on the Adult Civic Involvement data set has two records of data.

APPENDIX C

SAS CODE FOR DERIVED VARIABLES

This page is blank.

```

/*--COMMUNITY--*/

IF HCCOMMUN = 4 & HCCITY = 1 THEN COMMUNITY = 1;
    ELSE IF HCCOMMUN = 4 & HCCITY = 2 THEN COMMUNITY = 2;
    ELSE IF HCCOMMUN = 4 & HCCITY = 3 THEN COMMUNITY = 3;
    ELSE IF HCCOMMUN = 2 & HCSUB = 1 THEN COMMUNITY = 4;
    ELSE IF HCCOMMUN = 2 & HCSUB = 2 THEN COMMUNITY = 5;
    ELSE IF HCCOMMUN = 2 & HCSUB = 3 THEN COMMUNITY = 6;
    ELSE IF HCCOMMUN = 3 THEN COMMUNITY = 7;
    ELSE IF HCCOMMUN = 1 THEN COMMUNITY = 8;

/*--EMPLOYED--*/

IF (AWORK = 1 OR (AWORK = 2 AND ALEAVE = 1)) &
    AHOURS >= 35 THEN EMPLOYED = 1;
    ELSE IF (AWORK = 1 OR (AWORK = 2 AND ALEAVE = 1)) &
    AHOURS < 35 THEN EMPLOYED = 2;
    ELSE IF (AWORK = 2 AND ALEAVE = 2) &
    (APUBL = 1 OR APRIV = 1 OR AEMPL = 1 OR AREL = 1 OR
    AANSAD = 1) THEN EMPLOYED = 3;
    ELSE EMPLOYED = 4;

/*--HIGHGRAD--*/

DIPCNT = 0;
GRDIP = 0;
LE58 = 0;
EQ9 = 0;
GT9 = 0;

ARRAY HGRAD{16} GRADE1-GRADE16;
ARRAY SGRAD{16} SGRADE1-SGRADE16;
ARRAY HDIPL{16} SDIPL1-SDIPL16;
DO I = 1 TO 16;
    IF (HGRAD[I] > 9 OR SGRAD[I] = '17') THEN GT9 + 1;
    ELSE IF HGRAD[I] = 9 THEN EQ9 + 1;
    ELSE IF ((5 <= HGRAD[I] <= 8) OR ('15' <= SGRAD[I] <= '16'))
    THEN LE58 + 1;
    ELSE IF (HGRAD[I] = 4 OR HDIPL[I] = 1) THEN GRDIP + 1;
    ELSE IF (HDIPL[I] = 2 OR SGRAD[I] IN('N','K','T','P','1','2',
    '3','4','5','6','7','8','9','10','11','12','S','U'))
    THEN DIPCNT + 1;
END;

IF GT9 > 0 THEN HIGHGRAD = 5;
    ELSE IF EQ9 > 0 THEN HIGHGRAD = 4;
    ELSE IF LE58 > 0 THEN HIGHGRAD = 3;
    ELSE IF GRDIP > 0 THEN HIGHGRAD = 2;
    ELSE IF DIPCNT > 0 THEN HIGHGRAD = 1;

/*--RACEETHN--*/

IF HISPANIC = 1 THEN RACEETHN = 3;
    ELSE IF RACE = 1 THEN RACEETHN = 1;
    ELSE IF RACE = 2 THEN RACEETHN = 2;
    ELSE IF RACE IN(3,4) OR (RACE = 5 & OTHRAC IN(2,91))
    THEN RACEETHN = 4;

```

This page is blank.

APPENDIX E

DIRECTIONS AND CODE FOR LINKING DATA FILES

This page is blank.

Linking the NHES:96 Data Files

It is possible to link information from the four distinct NHES:96 files. This process is clear once the structure of the file identifiers is understood. First, the types of identifiers found on the NHES:96 data files are discussed. In the NHES:96, there are **household** identification numbers (**BASEID**), interview **subject** identification numbers (**ENUMID**), and **interview** or case identification numbers (**BASMID**). In addition, two-digit **person numbers** are provided for household members within households. The household, interview, and subject identification numbers are first discussed and then the person numbers.

- **BASEID**, is the **household** identification number. This eight-digit identification number is the same for every data record within a household and is the case identification number for the Household & Library data file. It is also provided on the Parent PFI/CI file, the Youth CI file, and the Adult CI file to permit data users to form linkages between the files.
- **ENUMID** is the **interview subject** identification number and is composed of 10 digits. "Interview subject" identification means that this number is unique to the person who is the subject of the interview. For example, in the Parent PFI/CI data file, **ENUMID** is the ID number of the child or youth who is the *subject* of the interview. Thus, **ENUMID** is the same in the Parent PFI/CI interview about a given youth and in that youth's own Youth CI interview record. **ENUMID** appears in the Parent PFI/CI and Youth CI data files only.
- **BASMID** is the unique **interview** or case identification number and is composed of 12 digits. Each Parent PFI/CI, Youth CI, or Adult CI interview has this unique interview ID. The first eight digits of **BASMID** are the same as **BASEID** for the household to which the interview belongs. The first 10 digits of **BASMID** are the same as the **ENUMID** of the subject of the interview. Therefore, a Parent PFI/CI interview record and a Youth CI record about the same youth would have the same value for **ENUMID**, but each interview would have its own unique **BASMID**. For the Parent PFI/CI interview, the last two digits are 01; for the Youth CI interview, the last two digits are 02; and for the Adult CI interview, the last two digits are 03. **BASMID** is the variable specified as the case identification number when creating a WesVarPC analysis from the Parent PFI/CI, Youth CI, or Adult CI files (see chapter 3 of this volume for a discussion of variance estimation and WesVarPC.)

Two-digit person identification numbers are provided on each of the NHES:96 data files in order to permit data users to copy information about certain individuals across interview records. In developing the public use data files, care was taken to include those person-level variables that were most likely to be needed by analysts. For example, the educational attainment of and languages spoken by the child's parents are included on the Parent PFI/CI file, since these parent characteristics are likely to be of interest to many analysts. These characteristics of subject adults are also included on the Adult CI file. In addition, since household characteristics (such as own/rent and income) are likely to be of interest to many analysts, these variables are contained on all four of the NHES:96 data sets and it is not necessary to copy them from the Household & Library file to the Parent PFI/CI, Youth CI, or Adult CI data files.

However, there may be circumstances in which an analyst would like to copy data items about a household member from one file to another. In order to facilitate linkages between the NHES:96 data files for the purpose of merging person characteristics, individual person numbers are provided on the data files. These two-digit numbers represent the number assigned to the person during enumeration of the household, i.e., 01 for the first person listed by the Screener respondent, 02 for the second person, etc.

- **PNUM(*n*)** is the two-digit person number variable in the Household & Library file. As noted above, many individual characteristics appear on this file for each household member, and these characteristics have sequential numbers, e.g., AGE1, AGE2, AGE3, and so on. Similarly, PNUM(*n*) carries the same sequential numbers in the variable name, so that it appears as PNUM1, PNUM2, PNUM3, etc. PNUM(*n*) contains the number assigned to each household member at the time of enumeration,

NOTE: *The value of PNUM(*n*) is not the same as the sequential number it carries in its variable name.* That is, PNUM1 is not necessarily equal to '01,' PNUM2 is not always equal to '02,' etc. There are two reasons for this. First, during data collection and data preparation, household members who were originally enumerated may have been deleted because they were later determined not to be household members according to the study definition. The enumeration numbers assigned to the household members were not changed when this occurred, because doing so would have disrupted linkages between segments in the hierarchical CATI data base. Second, after imputation of person-level records, the household members were sorted by age, oldest to youngest, before constructing the rectangular Household & Library file.

- In the Parent PFI/CI and Youth CI data files, four two-digit person numbers are provided -- **MOMNUM** for the child's mother (if any), **DADNUM** for the child's father (if any), **RESPNUM**, the person number of the respondent to the Parent PFI/CI interview, and **CHILDNUM** for the subject child or youth. If the mother or the father was the respondent to the Parent PFI/CI interview, MOMNUM (or DADNUM) will have the same value as RESPNUM.
- In the Adult CI data file, the person number for the sampled adult is contained in **PERSNUM**.

In order to effectively approach linkages between NHES:96 data files, it is important to remember the structure of the NHES:96 sample. Every household with a completed Screener interview has a household record in the Household & Library file. Therefore, every Parent PFI/CI, Youth CI, and Adult CI data record belongs to a household that is also represented in the Household & Library file. Because the Youth CI interview was only attempted if the corresponding Parent PFI/CI interview was completed, every Youth CI interview has a corresponding Parent PFI/CI record. As noted earlier, the sample of telephone numbers for the NHES:96 was split, and 95 percent was assigned to Parent/Youth interviewing and 5 percent was assigned to Adult CI interviewing. As a result, there are no Parent PFI/CI or Youth CI interviews for which there is an Adult CI interview in the same household.

The following examples are provided for the general populations for each component. Data users should consider the following tips regarding the length of time required to run a program and use of disk space:

- The data files are provided in ID order -- all of the following examples present code for sorting data files prior to linking (merging). Sorting the files can take up considerable time and disk space. If the files are already in the order required by the analyst, sorting is unnecessary.
- Keep only the variables required for your analysis -- specifying only the variables needed for the analysis will significantly improve the speed of the linking and the created data file will use less disk space. The use of a KEEP option, demonstrated in some of the following examples, can be used for this purpose.
- Keep only the relevant records -- when linking, for example, the Parent PFI/CI file with the Household & Library file, a match for every parent can be found that will bring together the Parent PFI/CI variables with the parents' related Household & Library variables. However, there are additional records in the Household & Library file unrelated to the Parent PFI/CI file, i.e., household information on respondents found in the Adult CI and Youth CI files. The example on **Linking between Parent PFI/CI and Youth CI files** demonstrates a technique for dropping unwanted records resulting from such a merge; in the SAS example note the use of the ONPARENT variable and in the SPSS examples the INPARENT recode.

Linking between Parent PFI/CI and Youth CI files is straightforward. The common identifier (key) is the ENUMID, and is available on both files. Sample SAS code to bring together these two files follows:

```
DATA TEMP;
MERGE parent_filename (IN=ONPARENT) youth_filename (IN=ONYOUTH);
BY ENUMID;
RUN;
```

Sample SPSS for DOS code is:

```
SET MORE = OFF

GET FILE = 'youth file'.
SORT CASES BY ENUMID.
SAVE OUTFILE = 'temp'.

GET FILE = 'parent file'.
SORT CASES BY ENUMID.

JOIN MATCH FILE = */FILE = 'temp'
  /BY ENUMID
  /MAP.
SAVE OUTFILE = 'merged file'.
```

Sample SPSS for Windows code is:

```
GET FILE = 'youth file'.
  /KEEP = ALL.
SORT CASES BY ENUMID.
SAVE OUTFILE = 'temp'.

GET FILE = 'parent file'.
  /KEEP = ALL.
SORT CASES BY ENUMID.

MATCH FILES FILE = */FILE = temp
  /BY ENUMID
```

```
/MAP.  
SAVE OUTFILE = 'merged file'.
```

Linking between the Household & Library file and the Parent PFI/CI, Youth CI, or Adult CI files requires using the key common to both files. This is accomplished using BASEID, which appears on all data files. The following example shows how to join selected library items from the Household & Library file with the Parent PFI/CI file and retain only records from the Parent PFI/CI file. Similarly, the Youth CI or Adult CI file can be substituted where references to the Parent PFI/CI file are made to allow joining library items with that file. Example SAS code to bring these files together follows:

```
DATA TEMP;  
MERGE parent_filename (IN=ONPARENT) household_and_library_filename (KEEP = BASEID LCOMP  
LCONSUME LDISTANC LJOBHELP);  
BY BASEID;  
IF ONPARENT;  
RUN;
```

Sample SPSS for DOS code is:

```
SET MORE = OFF  
  
GET FILE = 'household & library file'  
  /KEEP = BASEID LCOMP LCONSUME LDISTANC LJOBHELP.  
SORT CASES BY BASEID.  
SAVE OUTFILE = 'temp'.  
  
GET FILE = 'parent file'.  
COMPUTE INPARENT = 1.  
SORT CASES BY BASEID.  
SAVE OUTFILE = 'temp2'.  
  
JOIN MATCH TABLE='temp'  
  /FILE='temp2'  
  /BY BASEID  
  /MAP.  
  
SELECT IF (INPARENT = 1).  
SAVE OUTFILE = 'merged file'.
```

Sample SPSS for Windows code is:

```
GET FILE = 'household & library file'  
  /KEEP = BASEID LCOMP LCONSUME LDISTANC LJOBHELP.  
SORT CASES BY BASEID.  
SAVE OUTFILE = 'temp'.  
  
GET FILE = 'parent file'  
  /KEEP = ALL.  
COMPUTE INPARENT = 1.  
SORT CASES BY BASEID.  
SAVE OUTFILE = 'temp2'.  
  
MATCH FILES TABLE = 'temp'  
  /FILE='temp2'  
  /BY BASEID  
  /MAP.  
  
SELECT IF (INPARENT = 1).  
SAVE OUTFILE = 'merged file'.
```

Linking between selected household member characteristics and the Parent PFI/CI file

requires the use of household member person numbers. (The same approach can be used to link person-level variables in the Household & Library file with the Youth CI and Adult CI data files.) As noted above, the household member person numbers on the Parent PFI/CI file are stored in the variables CHILDNUM, MOMNUM, DADNUM, and RESPNUM. On the Household & Library file, the characteristics of household members have been stored in arrayed fields that number from 1 to 16, for example MARITL1-MARITL16 gives the marital status of each household member. Household member numbers have been stored in the fields PNUM1-PNUM16 on the Household & Library file. To determine the marital status of the child's father, each PNUM(*n*) field must be compared to the value of the DADNUM field and the corresponding arrayed MARITL(*n*) field contains the father's marital status, for example, if DADNUM equals the value in PNUM5, then the MARITL5 field contains the father's marital status. Building on the preceding code, the following code demonstrates a way this can be accomplished in SAS. Note that characteristics can be determined for the child, mother, or interview respondent by substituting CHILDNUM, MOMNUM, or RESPNUM for references to DADNUM in the following code.

```
DATA TEMP;
MERGE parent_filename (IN=ONPARENT) household_and_library_filename (KEEP = BASEID PNUM1-
PNUM16 MARITL1-MARITL16);
BY BASEID;
IF ONPARENT;
/* determine which element contains dads info and assign dads marital status to DADSMAR */
IF DADNUM = PNUM1 THEN DADSMAR = MARITL1;
ELSE IF DADNUM = PNUM2 THEN DADSMAR = MARITL2;
ELSE IF DADNUM = PNUM3 THEN DADSMAR = MARITL3;
ELSE IF DADNUM = PNUM4 THEN DADSMAR = MARITL4;
ELSE IF DADNUM = PNUM5 THEN DADSMAR = MARITL5;
ELSE IF DADNUM = PNUM6 THEN DADSMAR = MARITL6;
ELSE IF DADNUM = PNUM7 THEN DADSMAR = MARITL7;
ELSE IF DADNUM = PNUM8 THEN DADSMAR = MARITL8;
ELSE IF DADNUM = PNUM9 THEN DADSMAR = MARITL9;
ELSE IF DADNUM = PNUM10 THEN DADSMAR = MARITL10;
ELSE IF DADNUM = PNUM11 THEN DADSMAR = MARITL11;
ELSE IF DADNUM = PNUM12 THEN DADSMAR = MARITL12;
ELSE IF DADNUM = PNUM13 THEN DADSMAR = MARITL13;
ELSE IF DADNUM = PNUM14 THEN DADSMAR = MARITL14;
ELSE IF DADNUM = PNUM15 THEN DADSMAR = MARITL15;
ELSE IF DADNUM = PNUM16 THEN DADSMAR = MARITL16;
RUN;
```

Sample SPSS for DOS code is:

```
SET MORE = OFF
GET FILE = 'household & library file'
/KEEP = BASEID PNUM1 TO PNUM16 MARITL1 TO MARITL16.
SORT CASES BY BASEID.
SAVE OUTFILE = 'temp'.

GET FILE = 'parent file'.
COMPUTE INPARENT = 1.
SORT CASES BY BASEID.
SAVE OUTFILE = 'temp2'.

JOIN MATCH TABLE='temp'
/FILE='temp2'
/BY BASEID
/MAP.
SELECT IF (INPARENT = 1).

IF (DADNUM = PNUM1) DADSMAR = MARITL1.
IF (DADNUM = PNUM2) DADSMAR = MARITL2.
IF (DADNUM = PNUM3) DADSMAR = MARITL3.
IF (DADNUM = PNUM4) DADSMAR = MARITL4.
IF (DADNUM = PNUM5) DADSMAR = MARITL5.
```

```

IF (DADNUM = PNUM6) DADSMAR = MARITL6.
IF (DADNUM = PNUM7) DADSMAR = MARITL7.
IF (DADNUM = PNUM8) DADSMAR = MARITL8.
IF (DADNUM = PNUM9) DADSMAR = MARITL9.
IF (DADNUM = PNUM10) DADSMAR = MARITL10.
IF (DADNUM = PNUM11) DADSMAR = MARITL11.
IF (DADNUM = PNUM12) DADSMAR = MARITL12.
IF (DADNUM = PNUM13) DADSMAR = MARITL13.
IF (DADNUM = PNUM14) DADSMAR = MARITL14.
IF (DADNUM = PNUM15) DADSMAR = MARITL15.
IF (DADNUM = PNUM16) DADSMAR = MARITL16.

```

```

SAVE OUTFILE = 'merged file'.

```

Sample SPSS for Windows code is:

```

GET FILE = 'household & library file'
  /KEEP = BASEID PNUM1 TO PNUM16 MARITL1 TO MARITL16.
SORT CASES BY BASEID.
SAVE OUTFILE = 'temp'.

```

```

GET FILE = 'parent file'
  /KEEP = ALL.
COMPUTE INPARENT = 1.
SORT CASES BY BASEID.
SAVE OUTFILE = 'temp2'.
MATCH FILES TABLE = 'temp'
  /FILE='temp2'
  /BY BASEID
  /MAP.
SELECT IF (INPARENT = 1).

```

```

IF (DADNUM = PNUM1) DADSMAR = MARITL1.
IF (DADNUM = PNUM2) DADSMAR = MARITL2.
IF (DADNUM = PNUM3) DADSMAR = MARITL3.
IF (DADNUM = PNUM4) DADSMAR = MARITL4.
IF (DADNUM = PNUM5) DADSMAR = MARITL5.
IF (DADNUM = PNUM6) DADSMAR = MARITL6.
IF (DADNUM = PNUM7) DADSMAR = MARITL7.
IF (DADNUM = PNUM8) DADSMAR = MARITL8.
IF (DADNUM = PNUM9) DADSMAR = MARITL9.
IF (DADNUM = PNUM10) DADSMAR = MARITL10.
IF (DADNUM = PNUM11) DADSMAR = MARITL11.
IF (DADNUM = PNUM12) DADSMAR = MARITL12.
IF (DADNUM = PNUM13) DADSMAR = MARITL13.
IF (DADNUM = PNUM14) DADSMAR = MARITL14.
IF (DADNUM = PNUM15) DADSMAR = MARITL15.
IF (DADNUM = PNUM16) DADSMAR = MARITL16.

```

```

SAVE OUTFILE = 'merged file'.

```