

# **HEALTH AND RETIREMENT STUDY**

1996 Core

Final, Version 4.0

September 2003

Data Description and Usage



---

TABLE OF CONTENTS

---

<b>1.</b>	<b>OVERVIEW .....</b>	<b>1</b>
1A.	THE SAMPLE INTERVIEWED IN 1996.....	1
1B.	WHAT'S INCLUDED IN VERSION 4 .....	1
1C.	1996 QUESTIONNAIRE SECTIONS .....	1
1D.	LEVELS OF FILES IN VERSION 4 .....	2
<b>2.</b>	<b>FILE NAMING CONVENTIONS.....</b>	<b>3</b>
<b>3.</b>	<b>DATA FILES .....</b>	<b>3</b>
<b>4.</b>	<b>IDENTIFICATION VARIABLES.....</b>	<b>4</b>
4A.	PRIMARY IDENTIFICATION VARIABLES.....	4
4B.	PRIMARY IDENTIFICATION VARIABLES FOR DATASETS AT EACH OF THE SEVEN LEVELS .....	5
4C.	SECONDARY IDENTIFICATION VARIABLES .....	6
4D.	DATASETS INCLUDING ESUBHH AND OPN AS PRIMARY IDENTIFIERS.....	6
<b>5.</b>	<b>DISTRIBUTION FILES AND DIRECTORY STRUCTURE.....</b>	<b>7</b>
5A.	DISTRIBUTION FILES .....	7
5B.	DIRECTORY STRUCTURE .....	7
<b>6.</b>	<b>PROGRAM STATEMENTS .....</b>	<b>7</b>
6A.	USING THE FILES WITH SAS.....	8
6B.	USING THE FILES WITH SPSS .....	8
6C.	USING THE FILES WITH STATA .....	8
<b>7.</b>	<b>DOCUMENTATION .....</b>	<b>8</b>
7A.	CODEBOOK .....	9
7B.	OTHER TYPES OF DOCUMENTATION .....	15
<b>8.</b>	<b>ADDITIONAL NOTES .....</b>	<b>16</b>
8A.	BASIC DEMOGRAPHIC VARIABLES .....	16
8B.	HOUSEHOLDS WITH NO OR TWO FAMILY OR FINANCIAL RESPONDENTS .....	16
8C.	COMPARISON OF DATA IN VERSION 1, VERSION 2, VERSION 3, AND VERSION 4.....	16
8D.	UNFOLDING BRACKET VARIABLES AND IMPUTATIONS .....	17
8E.	MODULES.....	19
8F.	CALL-BACK INFORMATION .....	20
8G.	PARENT CHECKPOINTS .....	20
8H.	CHANGES IN DATA BECAUSE OF RESPONDENT COMMENTS.....	20
<b>9.</b>	<b>LONGITUDINAL ISSUES .....</b>	<b>21</b>
9A.	THE SAMPLE DESIGN.....	21

9B.	LINKING RESPONDENTS ACROSS TIME .....	21
9C.	LINKING OTHER PERSONS ACROSS TIME .....	21
9D.	CHARACTER TYPE IDENTIFICATION VARIABLES.....	22
9E.	IDENTIFICATION VARIABLE HHID HAS ONE EXTRA DIGIT .....	23
9F.	INAP CODES STORED AS BLANKS.....	24
<b>10.</b>	<b>OTHER DATA FILES .....</b>	<b>24</b>
10A.	TRACKER FILE .....	24
10B.	IMPUTATION FILE .....	24
10C.	RESPONDENT FILE WITH PARENT INFORMATION.....	24
<b>11.</b>	<b>OBTAINING THE DATA .....</b>	<b>25</b>
11A.	REGISTRATION AND DOWNLOADING THE DATA .....	25
11B.	CONDITIONS OF USE.....	25
11C.	PUBLICATIONS BASED ON HRS DATA.....	25
<b>12.</b>	<b>IF YOU NEED TO KNOW MORE .....</b>	<b>26</b>
12A.	HRS INTERNET SITE.....	26
12B.	CONTACT INFORMATION.....	26
<b>13.</b>	<b>APPENDIX .....</b>	<b>27</b>
13A.	EXAMPLES OF SUB-HOUSEHOLD AND RESPONDENT PERSON NUMBER AND OTHER PERSON NUMBER ASSIGNMENTS .....	27
13B.	CROSS-SECTIONAL MERGING.....	32

---

## 1. Overview

The 1996 HRS Core (Final, Version 4.0) data release includes data obtained as part of the Health and Retirement Study (HRS), a national longitudinal study that focuses on older Americans, and their health, retirement, and economic status. The HRS is a rich source of longitudinal, cross-sectional data for researchers and policymakers who study aging.

By receiving these data, which have been freely provided, you are agreeing to use them for research and statistical purposes only and make no effort to identify the respondents therein. In addition, you are in good faith agreeing to send us a copy of any publications you produce based on the data.

Funding has been provided by the National Institute on Aging at NIH, the Social Security Administration, the Department of Labor Pension and Welfare Benefits Administration, the Office of the Assistant Secretary for Planning and Evaluation at DHHS, the State of Florida Department of Elder Affairs, the NIH Office of Research on Minority Health, and the NIH Office of Research on Women's Health.

### 1A. The Sample Interviewed in 1996

The 1996 HRS Core (Final, Version 4.0) data release uses the HRS sub-sample. This sub-sample was interviewed previously in 1992 and 1994. In 1998, and every two years thereafter, the HRS and the three other sub-samples - AHEAD, War Babies, and Children of the Depression - were interviewed. For more details about the sub-samples, see [The Sample Design](#).

### 1B. What's Included in Version 4

In version 4, some variables previously included in the respondent-level coversheet file are now provided in a household-level coversheet file instead and a couple of background variables have been added to Section A. For more details about the versions, see [Comparison of Data in Version 1, Version 2, Version 3, and Version 4](#).

### 1C. 1996 Questionnaire Sections

The 1996 HRS data collection instrument or questionnaire contained the following sections.

- Section PR: Preload
- Section CS: CoverScreen
- Section A: Demographics
- Section B: Health Status
- Section C: Cognition
- Section PC: Proxy Cognition
- Section D: Family Structure
- Section E: Health Care Utilization
- Section F: Housing
- Section G: Employment (includes G, GG, GH and GD)
- Section H: Expectations
- Section J: Assets and Income
- Section N: Widowhood and Asset Changes
- Section R: Health Insurance
- Section M: Modules
- Section T: Other (Time Stamps, Thumbnails)

## **1D. Levels of Files in Version 4**

In the 1996 data collection instrument, most questions were asked of all respondents. Some questions were asked of a designated coversheet respondent or family respondent or financial respondent on behalf of the entire household.

In addition to the familiar household-level and respondent-level files, the 1996 HRS Core (Final, Version 4.0) data release contains files at five other levels - household-member-and-child, sibling, helper, transfer-to-child-or-grandchildren, and transfer-from-child.

### **1D1. Household-Level Files**

Household-level files contain questions that were asked about the household of a designated financial or family respondent. The financial respondent answered household-level financial questions on behalf of the entire household. Similarly, a family respondent answered family questions on behalf of the entire household. There were 6,816 households in which at least one interview was obtained in the 1996. The household-level files contain 6,816 records, one for each of these households.

### **1D2. Respondent-Level Files**

Respondent-level files contain questions that were asked of all respondents about themselves (or asked of a proxy about the respondent if the respondent was not able to answer for him or herself). The respondent-level files contain 10,964 records, one for each respondent or proxy who gave an interview in HRS 1996.

### **1D3. Household-Member-and-Child-Level Files**

These files contain information provided by the family respondent (sections PR, CS and D) or financial respondent (section J) about each household member or child of the respondent or the child of the respondent's spouse. The files contain 25,171 records, one for each household member or child.

### **1D4. Sibling-Level Files**

These files contain information provided by the family respondent about the respondent's siblings and siblings-in-law. The files contain 17,974 records, one for each sibling or sibling-in-law.

### **1D5. Helper-Level File**

This file contains information provided by each respondent about helpers other than the respondent's spouse or partner. A helper may be a person or organization who was reported by the respondent as providing help with ADLs or IADLs. If the helper is a child, the record contains information about the child and also about the child's spouse or partner, if any. The file contains 799 records, one for each helper or, if the helper is a married or partnered child, the helping couple.

### **1D6. Transfer-to-Child-or-Grandchildren-Level File**

This file contains information provided by the family respondent about transfers of money to a child and, collectively, children of a child. There were 3,442 reports of transfers to a child and 459 reports of transfers to children of a child. The combined transfer-to file contains 3,901 records, one for each transfer to a child or a child's children.

### 1D7. *Transfer-from-Child-Level-File*

This file contains information provided by the family respondent about transfers of money from children. The file contains 305 records, one for each transfer from a child.

## 2. File Naming Conventions

In the version 4 release, files are named beginning with "H96" for HRS 1996, next a letter or two designating the questionnaire section, next a separator, "\_" and finally one or two letters designating the level,

- H for household-level,
- R for respondent-level,
- MC for household-member-and-child-level,
- SB for sibling-level,
- HP for helper-level,
- TC for transfer-to-child-or-grandchildren-level, and
- FC for transfer-from-child-level.

Putting it all together, files named H96A\_R include variables from section A at the respondent-level.

The following extensions are used for the six different types of files that are distributed.

- .DA for data files,
- .SAS for SAS program statements,
- .SPS for SPSS program statements,
- .DO for Stata DO statements,
- .DCT for Stata dictionary statements, and
- .TXT for codebook files.

One of each of these file types is provide for each of 28 parts of the 1996 HRS Core (Final, Version 4.0) data release listed below. For example, H96A\_R.DA contains data, H96A\_R.SAS contains corresponding SAS program statements, H96A\_R.SPS contains corresponding SPSS program statements, H96A\_R.DO contains corresponding Stata DO statements, H96A\_R.DCT contains corresponding Stata dictionary statements, and H96A\_R.TXT contains the ASCII codebook.

## 3. Data Files

The 1996 HRS Core (Final, Version 4.0) data release is distributed in 28 parts. The parts are listed below along with the number of cases (N), number of variables (NV), and the primary identifiers (SORTBY).

### Household-level-files

H96CS_H	N=6816	NV=63	IDs= HHID ESUBHH
H96D_H	N=6816	NV=492	IDs= HHID ESUBHH
H96F_H	N=6816	NV=188	IDs= HHID ESUBHH
H96J_H	N=6816	NV=666	IDs= HHID ESUBHH
H96N_H	N=6816	NV=117	IDs= HHID ESUBHH

### Respondent-level-files

H96PR_R	N=10964	NV=97	IDs= HHID PN
H96CS_R	N=10964	NV=157	IDs= HHID PN

H96A_R	N=10964	NV=75	IDs=	HHID	PN
H96B_R	N=10964	NV=186	IDs=	HHID	PN
H96C_R	N=10964	NV=122	IDs=	HHID	PN
H96E_R	N=10964	NV=220	IDs=	HHID	PN
H96G_R	N=10964	NV=1006	IDs=	HHID	PN
H96H_R	N=10964	NV=39	IDs=	HHID	PN
H96J_R	N=10964	NV=49	IDs=	HHID	PN
H96M_R	N=10964	NV=542	IDs=	HHID	PN
H96N_R	N=10964	NV=76	IDs=	HHID	PN
H96PC_R	N=10964	NV=72	IDs=	HHID	PN
H96R_R	N=10964	NV=214	IDs=	HHID	PN
H96T_R	N=10964	NV=41	IDs=	HHID	PN

#### Helper-level-file

H96E_HP	N=799	NV=39	IDs=	HHID	PN	OPN
---------	-------	-------	------	------	----	-----

#### Household-member-and-child-level-files

H96PR_MC	N=25171	NV=25	IDs=	HHID	ESUBHH	OPN
H96CS_MC	N=25171	NV=16	IDs=	HHID	ESUBHH	OPN
H96D_MC	N=25171	NV=40	IDs=	HHID	ESUBHH	OPN
H96J_MC	N=25171	NV=26	IDs=	HHID	ESUBHH	OPN

#### Sibling-level-files

H96PR_SB	N=17974	NV=19	IDs=	HHID	ESUBHH	OPN
H96D_SB	N=17974	NV=29	IDs=	HHID	ESUBHH	OPN

#### Transfer-to-child-or-grandchildren-level-file

H96D_TC	N=3901	NV=22	IDs=	HHID	ESUBHH	OPN	ETYPTRAN
---------	--------	-------	------	------	--------	-----	----------

#### Transfer-from-child-level-file

H96D_FC	N=305	NV=22	IDs=	HHID	ESUBHH	OPN
---------	-------	-------	------	------	--------	-----

HRS 1996 data are provided in ASCII format, with fixed-length records. The records in the data files are sorted in order by the primary identifiers. You'll want to use associated SAS, SPSS or Stata program statements to read the data into the analysis package of your choice.

## 4. Identification Variables

Identification variables for HRS 1996 are stored in character format. For further information about merging 1996 data with prior years' data, see [Linking Respondents across Time](#) and [Linking Other Persons across Time](#).

### 4A. Primary Identification Variables

Five variables, HHID, ESUBHH, PN, OPN, ETYPTRAN are used in various combinations to uniquely identify the five different level datasets that comprise the version 4 data release.

#### 4A1. HHID -- Household Identifier

At the initial wave of data collection in 1992, each sample household was assigned a Household Identifier. The HHID is stable across waves of data



collection and uniquely identifies the original household and any households derived from that household in subsequent waves of data collection. HHID has six-digits.

#### **4A2. *ESUBHH -- HRS 1996 Sub-household Identifier***

In combination with HHID, ESUBHH uniquely identifies a household at the time of the HRS 1996 interview. Sub-household identifiers can be different at each wave. ESUBHH has one-digit. For more information, see [Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments](#).

#### **4A3. *PN -- Person Number***

In combination with HHID, PN uniquely identifies a respondent or respondent's spouse or partner. PNs are unique within an original household (HHID). The PN assigned to a particular respondent does not change across waves. PN has three-digits.

#### **4A4. *OPN -- Other Person Number***

In combination with HHID and ESUBHH, OPN uniquely identifies another person - household member, child, sibling, or helper. OPN has three-digits.

#### **4A5. *ETYPTRAN -- Type of Transfer***

In combination with HHID, ESUBHH, OPN, ETYPTRAN uniquely identifies a transfer. ETYPTRAN indicates whether the transfer was to a child or to the children of a child. ETYPTRAN has one-digit.

### **4B. Primary Identification Variables for Datasets at Each of the Seven Levels**

Two identifiers uniquely identify records in the respondent-level datasets:

- 1) HHID - HOUSEHOLD IDENTIFIER and
- 2) PN - PERSON NUMBER.

Three identifiers uniquely identify records in the helper-level datasets:

- 1) HHID - HOUSEHOLD IDENTIFIER,
- 2) PN - PERSON NUMBER and
- 3) OPN - OTHER PERSON NUMBER.

Two identifiers uniquely identify records in the household-level datasets:

- 1) HHID - HOUSEHOLD IDENTIFIER and
- 2) ESUBHH - 1996 SUB-HOUSEHOLD IDENTIFIER.

Three identifiers uniquely identify records in the household-member-and-child-level, sibling-level, and transfer-from-child-level datasets:

- 1) HHID - HOUSEHOLD IDENTIFIER,
- 2) ESUBHH - 1996 SUB-HOUSEHOLD IDENTIFIER, and
- 3) OPN - OTHER PERSON NUMBER.

Four identifiers uniquely identify records in the transfer-to-child-or-grandchildren-level datasets:

- 1) HHID - HOUSEHOLD IDENTIFIER,
- 2) ESUBHH - 1996 SUB-HOUSEHOLD IDENTIFIER,
- 3) OPN - OTHER PERSON NUMBER, and
- 4) ETYPTRAN - 1996 TYPE OF TRANSFER.

#### 4C. Secondary Identification Variables

In addition to the primarily identification variables that uniquely identify records in a dataset, secondary identification variables that allow links to other datasets are provided. Two sets of secondary identification variables are provided. They are listed below.

Secondary Identification Variables for  
respondent-level and  
helper-level datasets:

ESUBHH - 1996 SUB-HOUSEHOLD IDENTIFIER  
CSUBHH - 1994 SUB-HOUSEHOLD IDENTIFIER  
EPN\_SP - 1996 SPOUSE/PARTNER PERSON NUMBER  
ECSR - 1996 WHETHER COVERSHEET RESPONDENT  
EFAMR - 1996 WHETHER FAMILY RESPONDENT  
EFINR - 1996 WHETHER FINANCIAL RESPONDENT  
EQNR - 1996 QUESTIONNAIRE

Secondary Identification Variables for  
household-level,  
household-member-or-child-level,  
sibling-level,  
transfer- from-child-level, and  
transfer-to-child-or-grandchild-level datasets:

CSUBHH - 1994 SUB-HOUSEHOLD IDENTIFIER  
EPN\_SP - 1996 SPOUSE/PARTNER PERSON NUMBER  
EPN\_CS - 1996 COVERSHEET RESP PERSON NUMBER  
EPN\_FAM - 1996 FAMILY RESP PERSON NUMBER  
EPN\_FIN - 1996 FINANCIAL RESP PERSON NUMBER  
EPN\_NCS - 1996 NON-COVERSHEET RESP PERSON NUMBER  
EPN\_NFAM - 1996 NON-FAMILY RESP PERSON NUMBER  
EPN\_NFIN - 1996 NON-FINANCIAL RESP PERSON NUMBER  
EQNR\_CS - 1996 QUESTIONNAIRE - COVERSHEET RESP  
EQNR\_FAM - 1996 QUESTIONNAIRE - FAMILY RESP  
EQNR\_FIN - 1996 QUESTIONNAIRE - FINANCIAL RESP

#### 4D. Datasets Including ESUBHH and OPN as Primary Identifiers

When working with datasets including ESUBHH and OPN as primary identifiers -- household-member-and-child-level, sibling-level, transfer-from-child-level, or transfer-to-child-or-grandchildren-level datasets - it is essential to use ESUBHH to distinguish other persons.

Sometimes records with the same HHID and OPN will be separate reports about the same person. For example, in example 8 above where the couple with children divorces, the record where HHID is 89012, ESUBHH is 1, and OPN is 101 is the report about the child from the mother, and the record where HHID is 89012, ESUBHH is 2, and OPN is 101 is the report about the same child from the father.

At other times records with the same HHID and OPN will be reports about different persons. For example, in example 9 above where the couple divorces and both split-off households have new members, the record where HHID is 90123, ESUBHH is 1, and OPN is 151 is for the ex-wife's mother, while the record where HHID is 90123, ESUBHH is 2, and OPN is 151 is for the ex-husband's new stepchild. See also [Linking Other Persons across Time](#).

## 5. Distribution Files and Directory Structure

### 5A. Distribution Files

The files are packaged for download from our Web site in one big .ZIP file, H96CORE.ZIP, which contains seven smaller .ZIP files and this document, the data description, hrs96dd.pdf.

#### Data files

H96DA.ZIP containing data files.

#### Program statement files

H96SAS.ZIP containing SAS data descriptors.

H96SPS.ZIP containing SPSS data descriptors.

H96STA.ZIP containing Stata data descriptors.

#### Documentation files

H96CB.ZIP containing codebook.

H96QN.ZIP containing questionnaire files in .pdf format.

H96SPC.ZIP containing Surveycraft programming specifications.

You'll typically want to un-zip at least the data files, one set of program statement files, and the codebook files. The codebook files are also available on our Web site in a HTML format that includes an index of each section.

### 5B. Directory Structure

While a particular setup is not required for using HRS files, we have traditionally suggested a directory structure. By using this directory structure, you will not have to change the path name in your data descriptor files.

If you use a different structure, just change the directory references in the program files.

Directory	Contents
c:\hrs1996	H96core.zip
c:\hrs1996\data	Unzipped files from h96da.zip
c:\hrs1996\sas	Unzipped files from h96sas.zip
c:\hrs1996\spss	Unzipped files from h96sps.zip
c:\hrs1996\stata	Unzipped files from h96sta.zip
c:\hrs1996\codebook	Unzipped files from h96cb.zip
c:\hrs1996\qnaire	Unzipped files from h96qn.zip
c:\hrs1996\scraft	Unzipped files from h96spc.zip

Decompress the selected ZIP files into the appropriate subdirectories. You will need about 103 MB of free space on your storage device to store the 28 .DA files.

## 6. Program Statements

Each data file comes with associated SPSS, SAS or Stata program statements to read the data. Files containing SPSS statements are named with a .SPS extension,

those with SAS statements with a .SAS extension, and those with Stata statements with .DO and .DCT extensions.

The statement files are named beginning with the same prefix as the corresponding data file. For example, SAS statements in the file H96A\_R.SAS go with the H96A\_R.DA data file.

## 6A. Using the Files with SAS

To create a SAS system file for a particular dataset, two file types must be present for that dataset -- .SAS program statement files and .DA data files.

To create a SAS system file, load the \*.SAS file into the SAS Program Editor.

If the \*.SAS file is located in "c:\hrs1996\SAS" and the data file is located in "c:\hrs1996\DATA", you can run the file as is. A SAS system file (\*.SD2) will be saved to directory "c:\hrs1996\SAS".

If the files are not located in the specified directories, you will need to edit the \*.SAS file to reflect the proper path names prior to running the file.

## 6B. Using the Files with SPSS

To create an SPSS system file for a particular dataset, two file types must be present for that dataset -- .SPS program statement files and .DA data files.

To create an SPSS system file, open the \*.SPS file in SPSS as an SPSS Syntax File.

If the \*.SPS file is located in "c:\hrs1996\SPSS" and the data file is located in "c:\hrs1996\DATA", you can run the file as is. An SPSS system file (\*.SAV) will be saved to directory "c:\hrs1996\SPSS".

If the files are not located in the specified directories, you will need to edit the \*.SPS file to reflect the proper path names prior to running the file.

## 6C. Using the Files with Stata

To use Stata with a particular dataset, the following three file types must be present for that dataset -- .DCT files, .DO files, and .DA data files.

Files with the suffix "DA" contain the raw data for Stata to read. Files with the suffix "DCT" are Stata dictionaries used by Stata to describe the data. Files with the suffix "DO" are short Stata programs ("do files"), which you may use to read in the data. Load the .DO file into Stata and then submit it.

If the \*.DO and .DCT files are located in "c:\hrs1996\STATA" and the data file is located in "c:\hrs1996\DATA", you can run the .DO file as is.

If the files are not located in these directories, you must edit the \*.DO and \*.DCT files to reflect the proper path names before you run the files.

## 7. Documentation

There are several types of documentation available for use with the 1996 HRS Core (Final, Version 4.0) data release. In addition to this document, they include a codebook, programming specifications, and the box-and-arrow questionnaire.

## 7A. Codebook

The HRS 1996 Codebook is provided as a series of 28 ASCII text files and one master code text file. There is a codebook file corresponding to each data file. Each variable has its own codebook entry. Two examples are included below.

---

E769        B1. RATE HEALTH  
Section: B                    Level: Respondent            CAI Reference: Q769  
Type: Numeric                Width: 1                    Decimals: 0

B1. Next I have some questions about your health. Would you say your health is excellent, very good, good, fair, or poor?

.....  
1958            1. EXCELLENT  
3420            2. VERY GOOD  
3114            3. GOOD  
1686            4. FAIR  
781             5. POOR  
1               8. DK (don't know); NA (not ascertained)  
2               9. RF (refused)  
2               Blank. INAP (Inapplicable): [Q558:CS50] is (1)

---

E2246        F6.HOME VALUE \$  
Section: F                    Level: Household            CAI Reference: Q2246  
Type: Numeric                Width: 8                    Decimals: 0

F6. What is its present value? I mean, about what would it bring if it were sold today?

.....  
5                0. Actual value  
30               1-4999. Actual value (verified)  
4793            5000-500000. Actual value  
38               500001-99999996. Actual value (verified)  
1                99999997. Other  
432             99999998. DK (don't know); NA (not ascertained)  
36               99999999. RF (refused)  
1481            Blank. INAP (Inapplicable): [Q240:CS11] is (1) AND  
                 [Q420:CS26] is (NE 1); [Q377:R FIN-NON  
                 V106] is (NE 2); [Q2226:F3] is (7 OR DK OR  
                 RF); [Q2226:F3] is (2); [Q2229:F4] is (7 OR  
                 DK OR RF); [Q2229:F4] is (3); [Q2234:F5] is  
                 (7 OR DK OR RF); [Q2234:F5] is (4); partial  
                 interview; no financial respondent

---

### ASSIGNMENT STATEMENTS

- \* If [Q226:CS4] IS (1 OR 3) AND [Q253:CS15a] IS (NE 2) then [E341.R FAMILY/FINANCIAL TYPE] = value of [Q212:R W1 FAMFIN]
- \* If [Q233:CS9] IS (5) AND [Q234:CS10] IS (5) then [E341.R FAMILY/FINANCIAL TYPE] = value of [Q212:R W1 FAMFIN]
- \* If ([Q226:CS4] IS (5) OR [Q253:CS15a] IS (2)) AND [Q230:CS7] IS (NE 1) AND [Q231:CS8] IS (NE 1) then [E341.R FAMILY/FINANCIAL TYPE] = 3

```

* If [Q230:CS7] IS (1) OR [Q231:CS8] IS (1) OR [Q233:CS9] IS (1) OR
  [Q234:CS10] IS (1) then [E341.R FAMILY/FINANCIAL TYPE] = 3
* If [Q227:CS5] IS (5) then [E341.R FAMILY/FINANCIAL TYPE] = 3

```

```

-----
E341 R FAMILY/FINANCIAL TYPE
      Section: CS           Level: Household           CAI Reference: Q341
      Type: Numeric        Width: 1                 Decimals: 0

```

User note: This value was updated by Q347.

```

.....
1042      1. FINANCIAL R
1492      2. FAMILY R
3879      3. FINANCIAL & FAMILY R
403       4. NON-FINANCIAL & NON-FAMILY R
          Blank. INAP (Inapplicable); [Q220:WHICH IW 1ST R-2ND R] IS
          (NE 1)

```

### 7A1. Variable and Question Information

A horizontal line of equal signs sets off codebook entries from each other. Variable and question information is separated from codeframe information by a line of periods.

#### 7A1a. Assignment Statements

Assignment statements are included for variables that were assigned a value by the CAI program. No question may have been asked, or a question may have been asked only of respondents not assigned a value. If present, assignment statements are followed by a line of dashes and precede any other variable information. If more than one assignment statement is present, assignments are made by the first condition that was satisfied, i.e., they may be read as a series of IF-THEN-ELSE statements.

#### 7A1b. Variable Information

The variable information is contained on three lines. The first eight columns of the first line contain the variable name. Variable names that are specific to the 2000 HRS cross-section begin with an "E", i.e., " E769". Other variable naming conventions are described below. To the right of the variable name is the variable label.

The next two lines contain six fields with additional variable information.

#### Section

This is the section letter of the data files. This differs slightly from the letter of questionnaire section. For instance included in data file for section G are questions labeled G, GG, GH, and GD. Questionnaire section EV is found in data file section R.

#### Level

This indicates whether the data were collected at the household-level, respondent-level or another level.

```

"Household" for household-level files,
"Respondent" for respondent-level files,
"HHMemberChild" for household member/child-level files,

```

"Sibling" for sibling-level files,  
"ToChild" for transfer-to-child-or-grandchildren-level files,  
"FromChild" for transfer-from-child-level files, and  
"Helper" for helper-level files.

## CAI Reference

The CAI Computer Assisted Interviewing) Reference begins with the letter "Q". For data collected as part of the interview, the variable name and CAI name differ only in this first letter and in any extension to the variable name distinguishing multiple mention and looped variables. These numbers are in the range 1 to 5991. The 9000 series is used as the CAI Reference for identification variables. The 8200 series is used for a few derived variables. The 19000 series is used for masked variables.

## Type

Variable type -- numeric or character. Most variables are stored as numeric with the exception of the identification variables, which are stored as characters.

## Width

Variable width in the ASCII data file.

## Decimals

The number of implicit decimals. Usually zero.

## Question Text and User Note

This is followed, for most variables, by the question text and, for a few variables, by a user note containing supplemental information provided by HRS staff.

### 7A2. Codeframe Information

After a line of periods, codeframe information for each variable is provided.

#### 7A2a. Code Categories or Ranges

For categorical values single code values are provided. For continuous variables, ranges are designated in the codebook, e.g., "1992-1996"; however, actual, continuous values, e.g., 1992, 1993, 1994, etc., are provided in the data. Blanks are used throughout for INAP.

#### 7A2b. Frequencies

To the left of each code is its frequency. The frequencies are unweighted. They can be used to gain an initial overview of the data and to verify that data read into your statistical package accurately.

#### 7A2c. Code Text

In general, code text picked up from the data collection instrument specifications is in all capital letters. Code text not from the data collection instrument is in upper and lower case letters. These include codes added as the result of the comment review process. See [Changes in Data because of Respondent Comments](#) for further description of comment review process.

For values that were determined to be unlikely, the data collection instrument required verification of the value. These are designated in the codebook as "Actual value (verified)".

## INAP Code Text

Extensive documentation of the INAP category derived from the data collection instrument program specifications is provided. The rules specify the CAI reference and do not include loop or multiple mention information.

In a series of variables derived from multiple-mention question, the INAP category of mentions other than the first will include respondents who provided no second, third, etc. response even though the INAP text does not explicitly state this.

In the INAP text, 0 means a value of zero or blank or not answered. For instance, the phrase "[Q3796:H4] IS (00)" would include both a value of 0 and a value of blank for E3796, the variable with the CAI reference Q3796.

The rules in the INAP text do not reference codes added to the codeframe as a result of the comment review process. Nor do they reflect occasional, minor changes made to variable labels in processing.

Assignment statements may also return blanks; these blanks will not be documented in the INAP code text.

## Meaning of Certain Symbols

When used in the codebook, typically in the INAP text, the (A) symbol means "answered". For example, "[Q4217:J143c] is (A)" can be read as "if Q4217 was answered". The (NE "| ") symbol means "not blank". For example, "[Q2142:E158-2] is (NE "| ") can be read as "if Q2142 is not blank".

## 7A3. *Special Considerations*

### 7A3a. Variable Naming Conventions

Variable names that are specific to the HRS 1996 cross-section begin with an "E", i.e. "E769".

### Multiple-response and Looped Variables

Variables from questions that allowed multiple-mentions have a "M" in their name and a suffix that designates the mention, e.g., E3945M1, E3945M2, or E3945M3.

Variables from questions that were in a loop have an "\_" in their name and a suffix that designates the loop, e.g., E2840\_1, E2840\_2, or E2840\_3.

Variables from questions in loops that allowed multiple mentions end with a three-digit number indicating the loop and the mention. For example, for a question that was in a loop of two and allowed up to 10 mentions twenty variables would result, ending in 001-010 for the ten mentions in the first loop and 011-020 for the ten mentions in the second loop. Thus E5173001, E5173002, E5173003, etc. for mentions from the first loop and E5173011, E5173012, E5173013, etc. for mentions from the second loop.

### Masked Variables

We are dedicated to maintaining the confidentiality of our respondents. For that reason, a number of variables have been masked or are simply not included in the Final Release public dataset.

Names, addresses, days of birth, information on geographical relocation and similar variables are not included in publicly released files.



Geographical locations are recoded to a level no more detailed than U.S. Census Region and Division.

Data on the highest educational degree earned have been further grouped together to increase cell sizes.

Industry and occupation codes have been recoded into a limited number of categories from the original three-digit U.S. Census code.

The names of variables that were masked for confidentiality end in the letter "M"; for example, variable "E667M" is the masked version of variable "E667". In all, the following variables have been masked for confidentiality.

E640M	A2A. REGION - US BORN	H96A_R
E644M	A2B. REGION - NON-US BORN	H96A_R
E652M	A3C.R HIGH DEGREE - MASKED	H96A_R
E667M	A8.R RACE - MASKED	H96A_R
E715M	A27. REGION WHERE LIVE WHEN IN SCH-MASKE	H96A_R
E665M1M	A7A.TYPE HISPANIC- MASKED	H96A_R
E918M1M	B19A-1. MENTION HEALTH-1 - MASKED	H96B_R
E918M2M	B19A-1. MENTION HEALTH-2 - MASKED	H96B_R
E918M3M	B19A-1. MENTION HEALTH-3 - MASKED	H96B_R
E430M	CS33. REGION CURRENT RESIDENCE	H96CS_H
E434M	CS35. REGION OTHER RESIDENCE	H96CS_H
E419M	CS25B. REGION FACILITY WHERE	H96CS_R
E206M	CUR VQ VERSION	H96CS_R
E1590_1M	D106.STATE THEY LIVE - REGION	H96D_H
E1590_2M	D106.IN-LAWS STATE THEY LIVE - REGION	H96D_H
E1604_1M	D115.STATE MOTHER LIVE - REGION	H96D_H
E1604_2M	D115.STATE MOTHER-IN-LAW LIVE - REGION	H96D_H
E1620_1M	D125.STATE FATHER LIVE - REGION	H96D_H
E1620_2M	D125.STATE FATHER-IN-LAW LIVE - REGION	H96D_H
E1661M	D143C.SIB HIGHEST DEGREE-MASKED	H96D_SB
E2664M	G23. PREVIOUS WAVE WORK DONE	H96G_R
E2666M	G24/G25. CURRENT OCC	H96G_R
E2730M	G41a/G42. IND	H96G_R
E2732M	G41a/G42. OCC	H96G_R
E3131M	GG3a/GG6. IND	H96G_R
E3134M	GG3a/GG6. OCC	H96G_R
E3334M	GH5/GH6. IND	H96G_R
E3336M	GH5/GH6. OCC	H96G_R
E3508M1M	GD1A-1. CAUSE PROB-1 - MASKED	H96G_R
E3508M2M	GD1A-1. CAUSE PROB-2 - MASKED	H96G_R
E3508M3M	GD1A-1. CAUSE PROB-3 - MASKED	H96G_R
E5652M	M1-5. MOST SEVERE CONDITION - MASKED	H96M_R
E5653M1M	M1-5A.ANY OTHERS - MASKED	H96M_R
E5653M2M	M1-5A.ANY OTHERS - MASKED	H96M_R
E5654M	M1-6. TYPE OF WORK DAD DID	H96M_R

### Flag Variables

Much of Section A was asked only of new respondents. For your convenience, past wave data were merged into a few of these variables. The source of each observation can be determined using the associated variables ending in an "F". For example, variable "E638F" indicates where the data came from for each observation in variable "E638".

The following variables are flags that indicate the source of data for the associated variable.

E636F	A1. PAST WAVE MERGE FLAG - R BIRTH MO	H96A_R
E638F	A1B. PAST WAVE MERGE FLAG - R BIRTH YR	H96A_R
E649F	A3. PAST WAVE MERGE FLAG - R EDUC	H96A_R
E664F	A7. PAST WAVE MERGE FLAG - R HISPANIC	H96A_R
E667F	A8. PAST WAVE MERGE FLAG - R RACE	H96A_R

As described in [Parent Checkpoints](#), flag variables have also been provided to indicate whether a corrected value for the checkpoint was provided.

E1666_2F	D145.CKPT PARENTS-IN-LAW ALIVE-FLAG	H96D_H
E1557_2F	D90.MOTHER-IN-LAW LIVING-FLAG	H96D_H
E1566_2F	D93.FATHER-IN-LAW LIVING-FLAG	H96D_H

### Variables with Alternative Format

A few variables were derived. They are placed near an associated variable with an "A" extension.

E256A	MARRIAGE STATUS - CORRECTED	H96CS_R
E1666_2A	D145.CKPT PARENTS-IN-LAW ALIVE-CORRECTED	H96D_H
E1557_2A	D90.MOTHER-IN-LAW LIVING-CORRECTED	H96D_H
E1566_2A	D93.FATHER-IN-LAW LIVING-CORRECTED	H96D_H
E2120A	HELPER RELATIONSHIP COMBINED SOURCE	H96E_HP
E159A	SIB OR SIB-IN-LAW	H96PR_SB

### 7A3b. Banked-screen Questions

For some questions one question resulted in two variables. For instance, in Section G there are many questions that asked about amounts where the answers were given in the form of "amount" "per" period. For example, the question GG14 asking "How much were you earning, before taxes and other deductions, when you started?" results in two variables E3149 -- GG14.EARNINGS WHEN STARTED, and E3150 -- GG14.EARNINGS WHEN STARTED PER where E3149 is a dollar amount and E3150 is a time period. In the codebook you'll find the question text with the first of these two variables, E3149, and a line "AMOUNT: PER:". The second of these two variables, E3150, includes no question text in the codebook. See the previous variable, E3149, for the question text.

Similarly, in Section GH you'll find questions that asked about the date of an event that generated two variables, month and year. For example, the question GH29a asking "In what month and year did you start to receive these benefits?" followed in the codebook with "MONTH: YEAR:" generated two variables, E3413\_1 -

- GH29A.STARTED BENEFITS-MONTH and E3414\_1 -- GH29A.STARTED BENEFITS-YEAR. In the codebook the first variable of this pair includes the question text; the second does not.

### 7A3c. Two Files Combined to Create Transfer-to File

Two parallel sets of questions were asked about transfers to a child and transfers to the children of a child and a combined file was created. In the codebook you'll see two questions providing information for a single variable. And you'll see two distinct parts to the INAP code text documenting the INAP code for each of the component variables.

#### Transfer to child questions

D52S.COUNTER KID LOOP  
D53.TRANSFER TO CHILD \$ AMOUNT  
D53a.DK D13-1  
D53b.DK D13-2  
D53c. DK D13-3  
D53D.TRANSFER TO AMOUNT <500  
D54.NEXT CHILD TRANSFER TO

#### Transfer to grandchildren questions

D56S.COUNTER KID LOOP  
D57.TRANSFER TO OFFSPRING \$ AMOUNT  
D57a.DK D13-1  
D57b.DK D13-2  
D57c. DK D13-3  
D57D.TRANSFER TO AMOUNT <500  
D58.NEXT CHILD OFFSPRING

### 7A3d. Master Code

A master code file contains codeframes used in other sections of the codebook. The master codes include health conditions, health conditions - alphabetical list, occupation codes, industry codes, and state and country codeframes.

## 7B. Other Types of Documentation

In addition to this document and the codebook, two other types of documentation are available.

### 7B1. Programming Specifications

HRS 1996 data were collected using the Surveycraft CAI (Computer Assisted Interviewing) data collection software. Surveycraft software reads in programming specification files to determine what the software is supposed to do. The file "H96SPC.TXT" is a basic translation of the HRS 1996 Surveycraft program specifications. The TXT file is in ASCII format and can be viewed in a text editor.

This file retains much of the Surveycraft program language syntax and is, in places, not easy to decipher. However, it is the ultimate source for how the data were collected and the flow of the data collection instrument.

### 7B2. Box and Arrow Questionnaire

The research community has referred to the type of documentation that describes the questions asked in the interview as a "questionnaire". Since the 1996 HRS data were collected using a CAI program, a traditional hard-copy questionnaire was not produced as part of the data collection phase. Although, we have not created a questionnaire specifically for HRS 1996 Final Release, for your convenience, we are releasing the questionnaire that was created for the HRS 1996 Early Release to help document the asking sequence of the questions. These files are distributed in .PDF format and are available in the public file download area and on-line at our Web site. Section D (family structure) is not available.

## 8. Additional Notes

Found here are miscellaneous additional notes regarding 1996 HRS Core (Final, Version 4.0) data release. If we become aware of additional issues, they will be posted on our Web page in the Data Alerts section.

### 8A. Basic Demographic Variables

We have included many variables in the preload and coversheet sections because they determine the flow of the subsequent interview. Some of these variables were updated by information obtained later in the interview. These variables typically have "UPDATED" in their label. The codebook documentation for the variable includes a User Note indicating the variable that provided the updated information. However, all this can become confusing. In addition to the "F" variables listed above, [Flag Variables](#), for which we merged data from previous years, we recommend you use these variables for your analysis.

E256A	MARRIAGE STATUS - CORRECTED	H96CS_R
E374	R SEX	H96CS_R
E382	SP GENDER	H96CS_R
E10	HHMEM REL TO FAMILY R	H96PR_MC
E11	HHMEN REL TO NONFAM R	H96PR_MC

### 8B. Households with No or Two Family or Financial Respondents

As stated earlier in this document, the data collection design was to have asked most questions of all respondents, some questions of just a designated financial or family respondent on behalf of the household. However, occasionally that is not what happened. For some households we did not obtain an interview from a financial or family respondent. And for a few other households we obtained interviews with two financial or family respondents.

#### 8B1. Households with No Family or Financial Respondent

Although fifty-five households had no family respondent and sixty households had no financial respondent, household records are provided for these households with null values for the missing information. Households missing a family respondent can be identified by values of EPN\_FAM - 1996 FAMILY RESP PERSON NUMBER. Households missing a financial respondent can be identified by values of EPN\_FIN - 1996 FINANCIAL RESP PERSON NUMBER.

#### 8B2. Households with Two Family or Financial Respondents

In addition eleven households had two family or financial respondents; the responses of one of the two respondents were selected for inclusion in the household-level files. The respondents whose responses were used in the household records are indicated by the values of EFINR, 1996 WHETHER FINANCIAL RESPONDENT in the respondent record. The family respondents whose responses were used in the household records are indicated by the values of EFAMR, 1996 WHETHER FAMILY RESPONDENT.

### 8C. Comparison of Data in Version 1, Version 2, Version 3, and Version 4

Version 4 of the 1996 HRS Core Final Data Release incorporates a number of minor changes.

- o Some variables previously included in the respondent-level coversheet file are now provided in a household-level coversheet file instead.

- o A number of other variables were added. Included in version 4 are four variables from Section A that were not included in previous versions:
  - E709           A26a.HAVE YOU ALWAYS LIVED AT RES
  - E710           A26b. LIVE GEN AREA BEFORE
  - E711           A26c.YEARS LIVED MAIN HU AREA
  - E712           A26c.SINCE YEAR LIVE MAIN HU
- o Also added in version 4 are a number of variables that, while not of particular substantive interest, are references in assignment statements or in INAP text.

Retrospectively, major differences in Versions 1, 2, and 3 were:

- o Version 1 of the 1996 HRS Core (Final) data release included only sections CS, A, B, F, J, N and R.
- o Version 2 included all sections, but did not include variables identified as "family" variables most of which are located in sections D and J.
- o Version 3 included all sections and all variables.

In brief, the major differences in variables included in versions 1, 2, and 3 were:

Many household family variables added to section D in version 3.  
 Household financial variables added to section J in version 3.  
 Bracket components included in version 3.  
 Bracket summary variables not included in version 3.  
 "Standard" IDs included in version 2 and version 3.  
 Continue variables only included in version 2.  
 Timestamp variables only included in version 1.

See Detailed Listing of Variables in Versions 1, 2, 3, and 4 at our Web site for a complete listing variables not included in each of the four versions.

#### 8D. Unfolding Bracket Variables and Imputations

If a respondent did not give an exact report for a question, a series of questions was often asked that allowed us to bracket a response to within a certain range.

In versions 1 and 2 of the HRS 1996 Final Data Release, rather than providing all of the bracket questions, we provided a summary variable, which contained summary information. These summary variables had a "B" at the end of the variable name. For example, variable "E2247B" was a summary of bracket questions Q2247 through Q2250, which were removed from the publicly distributed data.

In versions 3 and 4 instead of creating these summary bracket variables, imputations for most unfolding bracket variables have been derived. The imputed variables are available in 1996 HRS Core Imputations (Final, Version 3.0) (December 13, 2002). See the section [Imputation File](#) for more details about the contents of this file. The unfolding bracket questions, or components, are included in the versions 3 and 4, 1996 HRS Final data release for all of the unfolding bracket sequences.

Imputations were not done for a few unfolding bracket sequences. Unfolding sequences for which imputations were not done are listed below. In the listing,

"Amount" is CAI Reference of the lead-in variable, if any

"Begin" is the CAI Reference of the first of the unfolding questions  
 "End" is the CAI Reference of the last of the unfolding questions  
 "File" is the level of the file where the variables are found  
 "Label" indicates the unfolding questions and the lead-in question

Amount	Begin	End	File	Label
---	Q894	895	R	B15B-C.INCONTINENCE DAYS
Q1471	Q1472	Q1475	TC	D57A-C.TRANSFER TO OFFSPRING \$ AMOUNT
Q1725	Q1726	Q1726	H	D151B-C.HELP TO FRIEND \$
Q1729	Q1730	Q1731	H	D152B-C.HELP FROM FRIEND \$
Q3828	Q3829	Q3832	R	H12B-E.RECEIVE INHERIT AMOUNT \$
Q4613	Q4614	Q4617	MC	J270A-D.(J67B2)\$ HHMEM EARNED LCY
Q4619	Q4620	Q4623	MC	J272A-D.(J67B2)SP \$ HHMEM EARNED LCY
Q4956	Q4957	Q4959	H	N30A-C.VALUE OF ASSETS
Q4960	Q4961	Q4963	H	N31A-C.VALUE OF DEBTS
Q4965	Q4966	Q4968	H	N33A-C.VALUE OF ASSETS
Q4969	Q4970	Q4972	H	N34A-C.VALUE OF DEBTS
---	Q5386	Q5388	R	M-ALL5-7.HOME VALUE IF SELL \$ <> 150K
---	Q5389	Q5391	R	M-ALL8-10.HOME VALUE IF SELL \$ <> 50K
---	Q5392	Q5394	R	M-ALL11-13.HOME VALUE IF SELL \$ <> 100K
---	Q5395	Q5397	R	M-ALL14-16.HOME VALUE IF SELL \$ > 150K
---	Q5398	Q5400	R	M-ALL17-19.HOME VALUE IF SELL \$ > 50K
---	Q5401	Q5403	R	M-ALL20-22.HOME VALUE IF SELL \$ > 100K
---	Q5404	Q5406	R	M-ALL23-25.HOME VALUE IF SELL \$ >= 150K
---	Q5407	Q5409	R	M-ALL26-28.HOME VALUE IF SELL \$ >= 50K
---	Q5410	Q5412	R	M-ALL29-31.HOME VALUE IF SELL \$ >= 100K
---	Q5417	Q5419	R	M-ALL34-36.CHECKING ACCOUNT \$ <> 25K
---	Q5420	Q5422	R	M-ALL37-39.CHECKING ACCOUNT \$ <> 1K
---	Q5423	Q5425	R	M-ALL40-42.CHECKING ACCOUNT \$ <> 10K
---	Q5426	Q5428	R	M-ALL43-45.CHECKING ACCOUNT \$ > 25K
---	Q5429	Q5431	R	M-ALL46-48.CHECKING ACCOUNT \$ > 1K
---	Q5432	Q5434	R	M-ALL49-51.CHECKING ACCOUNT \$ > 10K
---	Q5435	Q5437	R	M-ALL52-54.CHECKING ACCOUNT \$ >= 25K
---	Q5438	Q5440	R	M-ALL55-57.CHECKING ACCOUNT \$ >= 1K
---	Q5441	Q5443	R	M-ALL58-62.CHECKING ACCOUNT \$ >= 10K
---	Q5448	Q5450	R	M-ALL63-65.GESTIMATIONS MILES <> 50K
---	Q5451	Q5453	R	M-ALL66-68.GESTIMATIONS MILES <> 10K
---	Q5454	Q5456	R	M-ALL69-71.GESTIMATIONS MILES <> 25K
---	Q5457	Q5459	R	M-ALL72-74.GESTIMATIONS MILES > 50K
---	Q5460	Q5462	R	M-ALL85-87.GESTIMATIONS MILES > 10K
---	Q5463	Q5465	R	M-ALL88-90.GESTIMATIONS MILES > 25K
---	Q5466	Q5468	R	M-ALL91-93.GESTIMATIONS MILES >= 50K
---	Q5469	Q5471	R	M-ALL94-96.GESTIMATIONS MILES >= 10K
---	Q5472	Q5474	R	M-ALL97-99.GESTIMATIONS MILES >= 25K
---	Q5844	Q5848	R	M9-14A1-E1.INT 5% AND SAVE 10K \$
---	Q5850	Q5854	R	M9-14A2-E2.INT 5% AND SAVE 10K \$
---	Q5865	Q5867	R	M10-1A-1C.AMNT RETIREMENT SAVINGS \$

## 8E. Modules

Eleven modules were included in the 1996 questionnaire. They are described briefly below. See our Web site for more information about modules asked in other years.

Module All - Consumption and Anchoring: This module contains questions on food consumption, as well as questions designed to assess the degree to which the responses to unfolding brackets are affected by the level of the entry point (i.e. the "anchoring effect". To yield adequate sample sizes these questions were asked of all respondents.

Module 1 - Health During Childhood: This module contains question on the individual's health when they were growing up (from birth to age 16), on parental family composition and on their parental family's economic status.

Module 2 - Pedigree: This module contains questions on the individual's health pedigree, whether the individual's parents are still living, the cause of death if deceased, health status if living, and cause and age of death of any deceased siblings.

Module 3 - Personality: This is a brief personality inventory based on the respondent's rating of how closely each of 12 descriptive words fits the respondent.

Module 4 - Medicare Attitudes and Preferences: This module attempts to measure respondents' attitudes and preferences toward Medicare. It includes questions on whether the respondent would prefer various cash equivalent dollar payments to Medicare insurance. It is thus a form of contingent valuation.

Module 5 - Volunteerism and Time Use: This module includes questions on volunteerism and on time use. The respondent is asked how many hours he spent last year on ten types of volunteer activities ranging from help religious organizations to helping neighbors. Similarly the hours spent in eleven time-use activities ranging from television watching to reading are obtained.

Module 6 - Preference Parameters for Consumption, Saving and Labor Supply: This module attempts to understand the respondents' preferences toward consumption, saving and labor supply by asking about their behavior if they won a hypothetical sweepstakes which would pay them an amount equal to their current family income for life.

Module 7 - Advanced Directives: This module contains questions on advanced directives and respondents preferences toward a hypothetical cancer treatment under a variety of costs and treatment success rate assumptions.

Module 8 - Attitudes toward Inter-Familial Transfers: This module contains questions on the respondents' willingness to give financial help to parents and/or siblings under a variety of hypothetical situations.

Module 9 - Retirement Planning: This module asks about retirement planning and saving for retirement and contains a question intended to measure the extent to which the respondent understands compound interest. It asks about the extent to which the respondent relies or plans to rely on Social

Security, employer provided pension plans, IRAs 401k or Keogh plans and private savings. For retirees it contains questions on the adequacy of savings.

Module 10 - Saving for Retirement: Contains questions on the current level of savings for retirement and on the reliance or expect reliance on public, private and personal sources of income in retirement.

## 8F. Call-Back Information

Partway through the data collection period, a data collection instrument programming error was discovered that meant that for some cases information about parents was not being collected. The error was corrected and respondents that should have been asked these questions were called back. This call-back information was then merged into the file with the original information.

For two respondents, we only have the call-back information, not the original information. These two cases can be identified by the value of E558 CS50.PARENTS SENDBACK.

## 8G. Parent Checkpoints

Two checkpoints were used to determine whether to ask follow-up questions about help given to parents or parents-in-law of the family respondent. The checkpoint for parents-in-law was E1666\_2, D145.CKPT PARENTS-IN-LAW ALIVE. For previously interviewed respondents who had a new spouse or partner as of the 1996 wave, and whose spouse or partner, if any, at the time of the previous interview was not dead, the variable indicates whether the parents of their current spouse or partner are alive at the time of the 1996 interview; for all other previously interviewed respondents who had a spouse or partner at the time of the 1996 interview, the variable indicates whether the parents of their spouse or partner at time of the previous interview were alive at the time of that interview.

However, the values in this checkpoint were not assigned correctly. See E1666\_2A for more correct values. See E1666\_2F for a flag indicating whether the corrected value is different than the original value. Similarly corrected values and flags are provided for E1557\_2, D90.MOTHER-IN-LAW LIVING and E1566\_2, D93.FATHER-IN-LAW LIVING.

The checkpoints were used to determine asking of follow-up questions. The values of E1666\_2 were used in the asking sequences of D145, D146 D147, D148, D149, and D150. The values of the follow-up questions have not been corrected.

## 8H. Changes in Data because of Respondent Comments

The interviewer recorded comments the respondent made during the course of the interview. HRS staff reviewed these comments for selected questions. If it was determined that the comment changed the substance of the recorded answer, the coded answer was changed. When such a code change was made, immediate subsequent questions that reference the changed question within the section were made consistent, but no consistency checking was done between sections. If these changes affected a branch-point question, the branch-point was not changed. Questions that reference the branch-point may be inconsistent with the earlier change.



No changes due to respondent comments were made to variables E5162-E5200 in section R due to a processing error. In Section J, double reporting of income or assets was corrected in the comment coding process wherever possible.

## 9. Longitudinal Issues

### 9A. The Sample Design

As of 1998, the HRS sample was comprised of four sub-samples.

The HRS sub-sample consists of people who were born 1931 through 1941 and were household residents of the conterminous U.S. in the spring 1992; and of their spouses or partners at the time of the initial interview in 1992. The HRS sub-sample was interviewed in 1992, 1994, 1996, 1998 and every two years thereafter.

The AHEAD sub-sample consists of people who were born in 1923 or earlier, were household residents of the conterminous U.S. in the spring 1992, and were still household residents at the time of their first interview in 1993 or 1994; and of their spouses or partners at the time of the initial interview. The AHEAD sub-sample was interviewed in 1993-94, 1995-96, 1998 and every two years thereafter.

The War Babies sub-sample consists of people who were born 1942 through 1947, were household residents of the conterminous U.S. in the spring 1992, and were still household residents at the time of the first interview in 1998; and their of spouses or partners at the time of the initial interview. The War Baby sub-sample was interviewed in 1998 and every two years thereafter.

The Children of the Depression sub-sample consists of people who were born 1924 through 1930, were household residents of the conterminous U.S. when first interviewed in 1998, and who, at that time, did not have a spouse or partner who was born before 1924 or between 1931 and 1947. The Children of the Depression sub-sample was interviewed in 1998 and every two years thereafter.

Original sample members are those selected as described above and their spouses or partners at the time of the initial interview in 1992 (HRS), 1993 (AHEAD) or 1998 (CODA or WB). For more details about the sample, see our Web site.

### 9B. Linking Respondents across Time

Respondent records from 1992 and 1994 may be linked to respondent records from 1996 by HHID and PN. The sub-household identifiers, ASUBHH for 1992, W2SUBHH for 1994 and ESUBHH for 1996 can be used to link cross-sectional household data with the cross-sectional respondent data. See [Character Type Identification Variables](#) for information about storage type of identification variables for merging 1992 and 1994 files with files from later years.

### 9C. Linking Other Persons across Time

Linking other person records across time is more complex. You can use the previous wave's SUBHH to link one wave's other person record with the corresponding other person record from the previous wave. For example, use HHID, CSUBHH, and OPN to link 1996 other person records with 1994 other person records. See our Web site for more detailed information.

For some 1996 other person records this will result in two 1996 records matching one 1994 record. In the example [Couple with Children and Siblings Divorces.](#) where the couple with children divorces, the 1996 record where HHID is 89012, ESUBHH is 1, and OPN is 101 is the report about the child from the mother and the record where HHID is 89012, ESUBHH is 2, and OPN is 101 is the report about the same child from the father. The value of CSUBHH for each of these 1996 records would 0. Each of these 1996 records would match the 1994 record about the child before the couple divorced as the 1994 record would have a HHID of 89012, a CSUBHH of 0 and a OPN of 101.

For other 1996 other person records this will result in no 1996 record matching a 1994 record. In the example [Couple Divorces, One Respondent Remarries, Both Split-off Households Have New Members.](#) where the couple divorces and both split-off households have new members, the record where HHID is 90123, ESUBHH is 1, and OPN is 151 is for the ex-wife's cohabiting mother, while the record where HHID is 90123, ESUBHH is 2, and OPN is 151 is for the ex-husband's new stepchild. The value of CSUBHH for each of these 1996 records would 0. However neither of these 1996 records would match a 1994 record before the couple divorced because no 1994 record would have a HHID of 89012, a CSUBHH of 0 and an OPN of 151.

However, be aware that the technique of matching OPN records by HHID, previous wave SUBHH and OPN does not work well for 1994 when the value of 999 was used for many OPNs, for persons who assumed the OPN number of their deceased spouse, for persons with more than one OPN or for OPNs used by more than one person.

## 9D. Character Type Identification Variables

In some prior releases of HRS, all variables were stored in NUMERIC format. Releasing identification variables in numeric format caused some problems. Thus starting with 1995 data collections, we are releasing all primary and secondary identification variables in character format and are no longer creating combined identification variables. As of this writing, only the 1992 and 1994 releases still include numeric identification variables.

Also released in character mode are variables that include an other person number that may be used for merging with a primary or secondary identification variable. Typically these are variables from questions asking who or which person did something. Examples are:

```
E1402    D16.CHILD CLOSEST TO MAIN RESIDENCE
E1583_1  D101a.WHICH SIBLING
E2056    E99.IADLS-WHO HELP,1
```

Most software packages will require that the identification variables in each dataset be of the same type. We suggest that analysts of 1992 and 1994 HRS datasets convert the numeric identifiers to character type. Each software package has different ways of converting variable types in an existing dataset. We provide examples of numeric to character/string conversion below.

In SAS, for example, to convert numeric identification variables to character variables, you can use code like this.

```
data out.w2a;
  *      place ATTRIB statement before SET statement to
        control order of variables in output file;
  attrib
```

```

CHHID label='HOUSEHOLD IDENTIFIER' length=$6 format=$char6.
CPN label='PERSON NUMBER' length=$3 format=$char3.;
set in.w2a;
*      numeric to character conversion;
CHHID=put(HHID,z6.0);
CPN=put(PN,z3.0);
drop HHID PN;
rename CHHID=HHID CPN=PN;
run;

```

In STATA to convert numeric identification variables to string variables, you can use code like this.

```

use w2a
*      numeric to string conversion
gen str6 chhid=substr(string(HHID+1000000),2,6)
gen str3 cpn=substr(string(PN+1000),2,3)
drop HHID PN
rename chhid HHID
rename cpn PN
lab var HHID "HOUSEHOLD IDENTIFIER"
lab var PN "PERSON NUMBER"
order HHID PN
sav, replace

```

In SPSS to convert numeric identification variables to string variables, you can use code like this.

```

get file 'c:\hrs1994\spssfiles\W2a.sav'.
string chhid (a6) cpn(a3).
variable labels chhid 'HOUSEHOLD IDENTIFIER' cpn 'PERSON NUMBER'.
compute chhid=string(hhid,n6).
compute cpn=string(pn,n3).
save outfile='c:\hrs1994\spssfiles2\w2a.sav'
  / drop=hhid pn
  / rename=(chhid=hhid) (cpn=pn)
  / keep hhid pn all
  / map.

```

Alternatively, before reading in datasets, you can change the variable type in the program statements themselves.

Package	Type	Statement for variable V1
SAS	numeric	V1 1-4
	character	V1 \$ 1-4
Stata	numeric	_column(1) byte V1 %4.0f "LABEL"
	character	_column(1) str4 V1 %4s "LABEL"
SPSS	numeric	V1 1-4
	character	V1 1-4 (A)

## 9E. Identification Variable HHID Has One Extra Digit

A "0" was added to the front of the identification variable HHID for HRS subsample households, in order to make it possible to combine HRS and AHEAD sub-

samples for the 1998 data collection. The change makes the old HRS five digit HHID six digits long. For example, old HHID "12345" becomes "012345" as of HRS 1996 and subsequent years. You will want to be aware of these changes when reading the 1992 and 1994 documentation for preparing merges with later years' data. Note that however, that the 1992 and 1994 data currently available on our web site have the longer, six-digit, HHID values.

## 9F. INAP Codes Stored as Blanks

Beginning with HRS 1996 data, INAP codes are stored in the ASCII data files as blanks. In previous waves, 1992 and 1994, INAPs had been assigned to a number, usually zero or a value ending in "96". If you are merging the 1996 data with data from 1992 or 1994, you will want to be aware of this difference.

## 10. Other Data Files

### 10A. Tracker File

The Cross-Wave: Tracker File contains a record for every individual who was interviewed one or more times during the period 1992 through the most recent interviewing year as well as a record for the non-responding spouse or partner of such a person. It provides useful information about a respondent's participation, or lack of participation, in all waves of the study.

In addition to the primary identification variables HHID and PN, the tracker file contains basic demographic information, and basic information about the interview year such as whether a person gave an interview in a particular wave as well as some basic information obtained from the National Death Index (NDI).

The tracker file also contains cross-sectional weights for each interview year 1992 through the most recent year for both respondent and household level analysis.

See our Web site for the most recent version of this file

### 10B. Imputation File

The 1996 HRS Core Imputations files are available at our Web site. In them you will find imputed values for the all unfolding bracket variables except those listed in the section about [Unfolding Bracket Variables and Imputations](#). Imputations are provided for the variables satisfying our ad hoc selection rules listed below.

- 1) The Bracket Rule: With the exceptions noted above, all the variables with brackets were imputed, and included in the imputation file.
- 2) The Income/Asset Component Rule: A variable without brackets were imputed if it was determined to be an integral component of household income, housing equity, or net wealth.
- 3) The Imputability Rule: A variable eligible for imputation based on Rule 1 or 2 was not imputed if there are not enough valid observations.

Also included are four aggregate variables -- main home equity, and second home equity, total assets, and household income.

### 10C. Respondent File with Parent Information

Information about the parents and parents-in-law of the family respondent was asked of the family respondent and is contained in the household-level H96D\_H

file. A special file has been prepared with the information about the family respondent's parents merged onto the family respondent's record and about the family respondent's spouse's or partner's parents merged onto the family respondent's spouse's or partner's record. This file is available at our Web site.

## 11. Obtaining the Data

### 11A. Registration and Downloading the Data

HRS data are available for free to researchers and analysts at the HRS Web site. In order to obtain public release data, you must first register at our Web site. Once you have completed the registration process, your username and password will be sent to you via e-mail. Your username and password are required to download any data files.

By registering all users, we are able to document for our sponsors the size and diversity of our user community allowing us to continue to collect these important data. Registered users receive user support, information related to errors in the data, future releases, workshops, and publication lists. The information you provide will not be used for any commercial use, and will not be redistributed to third parties.

### 11B. Conditions of Use

By registering, you agree to the Conditions of Use governing access to Health and Retirement public release data. You must agree to

- o not attempt to identify respondents
- o not transfer data to third parties except as specified
- o not share your username and password
- o include specified citations in work based on HRS data
- o provide information to us about publications based on HRS data
- o report apparent errors in the HRS data or documentation files
- o notify us of changes in your contact information

For more information concerning privacy issues and conditions of use, please read "Conditions of Use for Public Data Files" and "Privacy and Security Notice" at the Public File Download Area of the HRS Web site.

### 11C. Publications Based on HRS Data

As part of the data registration process, you agree to include specified citations and to inform HRS of any papers, publications, or presentations based on HRS data. Please send a copy of any publications you produce based on HRS data, with a bibliographical reference, if appropriate, to the address below.

Health and Retirement Study  
Attn: Papers and Publications  
The Institute for Social Research, Room 3050  
P.O. Box 1248  
Ann Arbor, MI (USA) 48106-1248

Alternately, you may contact us by e-mail to: [hqsquest@isr.umich.edu](mailto:hqsquest@isr.umich.edu) with "Attn: Papers and Publications" in the subject line.

## 12. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the HRS public data. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

### 12A. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site. Our URL (new, as of 9/1/2002) is:

<http://hrsonline.isr.umich.edu/>

### 12B. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

Internet: Help Desk at our Web site

E-mailto:[hqsquest@isr.umich.edu](mailto:hqsquest@isr.umich.edu)

Postal service:

Health and Retirement Study  
The Institute for Social Research, Room 3050  
The University of Michigan  
P.O. Box 1248  
Ann Arbor, MI 48106-1248

FAX: (734) 647-1186

## 13. Appendix

This appendix provides detailed examples of sub-household assignments and several examples of cross-sectional merging data files. An understanding of sub-household assignments is useful when merging files from different levels or different years. In addition to an overview, actual code of three commonly used languages, SAS, Stata, and SPSS, is provided in the section of merging data files.

### 13A. Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments

In the first year of data collection, all households, consisting of either a single respondent or of two married or partnered respondents, were assigned a SUBHH of 0.

In subsequent waves, a SUBHH of 0 indicates that the original household has not split due to divorce or separation of spouses or partners, although one member of a couple may have died or a single respondent may have become married or partnered.

A value of 1 or 2 indicates a household in which the original couple split, divorced or separated. One of the original couple is assigned a SUBHH of 1; the other is assigned a SUBHH of 2.

A value of 5 or 6 indicates a previously split household split a second time. One of the couple from a SUBHH 1 or 2 retains a SUBHH of 1 or 2; the other is assigned a SUBHH 5 or 6.

A value of 7 indicates respondents from split household reunited<sup>1</sup>.

It is important to understand these assignments when you merge records from different waves of the study.

#### 13A1. *Married Couple Stays Married.*

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 012345 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

At the time of the second cross-section the two respondents are still married, and each retains their HHID of 012345 and their SUBHH of 0 and his and her PN of 010 and 020, respectively.

Time 1

Household records

HHID=012345 ASUBHH=0

Respondent records

HHID=012345 PN=010 ASUBHH=0

HHID=012345 PN=020 ASUBHH=0

Time 2

---

<sup>1</sup> In addition, a SUBHH of 3 or 4 indicates the "household" a deceased respondent who is considered to be in a household of his or her own. These values do not occur in these files because all records in these files are from living respondents.

Household records  
HHID=012345 CSUBHH=0  
Respondent records  
HHID=012345 PN=010 CSUBHH=0  
HHID=012345 PN=020 CSUBHH=0

**13A2. Couple Divorces.**

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 023456 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 023456, but one is assigned a SUBHH of 1 and the other is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively.

Time 1  
Household records  
HHID=023456 ASUBHH=0  
Respondent records  
HHID=023456 PN=010 ASUBHH=0  
HHID=023456 PN=020 ASUBHH=0

Time 2  
Household records  
HHID=023456 CSUBHH=1  
HHID=023456 CSUBHH=2  
Respondent records  
HHID=023456 PN=010 CSUBHH=1  
HHID=023456 PN=020 CSUBHH=2

**13A3. One or Both Respondents Die.**

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 034567 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

One respondent dies before the next wave. At the next wave, both respondents retain their HHID of 034567. The living respondent retains her SUBHH of 0; the deceased respondent is assigned a SUBHH of 3. (If both respondents die, one would be assigned a SUBHH of 3 and the other would be assigned a SUBHH of 4.) Each original respondent retains his and her PN of 010 and 020, respectively.

Time 1  
Household records  
HHID=034567 ASUBHH=0  
Respondent records  
HHID=034567 PN=010 ASUBHH=0  
HHID=034567 PN=020 ASUBHH=0  
  
Time 2  
Household records  
HHID=034567 CSUBHH=0  
HHID=034567 CSUBHH=3 (in exit interview)  
Respondent records  
HHID=034567 PN=010 CSUBHH=3 (in exit interview)



HHID=034567 PN=020 CSUBHH=0

**13A4. Single Respondent Marries.**

A respondent who has never been married is in the first cross-section. The respondent is assigned a HHID of 045678 and a SUBHH of 0 and a PN of 010.

At the time of the second cross-section, the respondent has married. Both the respondent and her new spouse are assigned a HHID of 045678 and a SUBHH of 0 because the household was not divided. The original respondent retains her PN of 010. Her new spouse is assigned PN of 011.

Time 1

Household records

HHID=045678 ASUBHH=0

Respondent records

HHID=045678 PN=010 ASUBHH=0

Time 2

Household records

HHID=045678 CSUBHH=0

Respondent records

HHID=045678 PN=010 CSUBHH=0

HHID=045678 PN=011 CSUBHH=0

**13A5. Couple Divorces, One Respondent Remarries and Divorces.**

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 056789 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced and he has remarried. Both original respondents retain the HHID of 056789, but she is assigned a SUBHH of 1 and he is assigned a SUBHH of 2. His new spouse is also assigned the HHID of 056789 and the SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. His new spouse is assigned PN of 011.

By the time of the third cross-section, that new couple has gotten divorced. All respondents retain the HHID of 056789. The original sample member ex-wife has the SUBHH of 1. The original sample member ex-husband has a SUBHH of 2, and the non-original sample member, his second ex-wife, is assigned the SUBHH of 5. Each original respondent retains his and her PN of 010 and 020, respectively. His second ex-wife retains her PN of 011.

Time 1

Household records

HHID=056789 ASUBHH=0

Respondent records

HHID=056789 PN=010 ASUBHH=0

HHID=056789 PN=020 ASUBHH=0

Time 2

Household records

HHID=056789 CSUBHH=1

HHID=056789 CSUBHH=2

Respondent records

HHID=056789 PN=010 CSUBHH=2

HHID=056789 PN=011 CSUBHH=2  
HHID=056789 PN=020 CSUBHH=1

Time 3

Household records

HHID=056789 FSUBHH=1  
HHID=056789 FSUBHH=2  
HHID=056789 FSUBHH=5

Respondent records

HHID=056789 PN=010 FSUBHH=2  
HHID=056789 PN=011 FSUBHH=5  
HHID=056789 PN=020 FSUBHH=1

**13A6. Couple Divorces and Marries Again.**

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 067890 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 067890, but one is assigned a SUBHH of 1 and the other is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively.

By the time of the third cross-section, the respondents have remarried each other. Both are assigned the HHID of 067890 and the SUBHH of 7. Each original respondent retains his and her PN of 010 and 020, respectively.

Time 1

Household records

HHID=067890 ASUBHH=0

Respondent records

HHID=067890 PN=010 ASUBHH=0  
HHID=067890 PN=020 ASUBHH=0

Time 2

Household records

HHID=067890 CSUBHH=1  
HHID=067890 CSUBHH=2

Respondent records

HHID=067890 PN=010 CSUBHH=1  
HHID=067890 PN=020 CSUBHH=2

Time 3

Household records

HHID=067890 FSUBHH=7

Respondent records

HHID=067890 PN=010 FSUBHH=7  
HHID=067890 PN=020 FSUBHH=7

**13A7. Married Couple with Children and Siblings.**

At the time of the first cross-section, sample household with a HHID of 078901 contains two respondents assigned PNs of 010 and 020, respectively. Associated with the household are three children with OPNs of 101, 102, and 103, and two siblings with OPNs of 051 and 052. All seven persons will keep those same PNs and OPNs across time. A friend who lives with the respondents is assigned an

OPN of 080. The friend will keep her OPN of 080 across time only if she is a household member at each wave.

Time 1

Household records

HHID=078901 ASUBHH=0

Respondent records

HHID=078901 PN=010 ASUBHH=0

HHID=078901 PN=020 ASUBHH=0

Household member/child records

HHID=078901 ASUBHH=0 OPN=101 (child)

HHID=078901 ASUBHH=0 OPN=102 (child)

HHID=078901 ASUBHH=0 OPN=103 (child)

HHID=078901 ASUBHH=0 OPN=051 (sibling)

HHID=078901 ASUBHH=0 OPN=052 (sibling)

HHID=078901 ASUBHH=0 OPN=080 (friend)

### **13A8. Couple with Children and Siblings Divorces.**

At the time of the first cross-section, sample household with a HHID of 089012 contains two respondents assigned PNs of 010 and 020, respectively. Associated with the household are two children with OPNs of 101, and 102, and three siblings, her two brothers with OPNs of 051 and 052 and his sister with an OPN of 061. A friend who lives with the respondents is assigned an OPN of 080.

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 089012, but he is assigned a SUBHH of 2 and she is assigned a SUBHH of 1. Each original respondent retains his and her PN of 010 and 020, respectively.

The two children appear both in their father's SUBHH 2 and also in their mother's SUBHH 1 with their respective OPNs, 101, and 102. The three siblings appear in the SUBHH of their respective siblings and maintain their respective OPNs. The ex-wife's brothers appear as part of her SUBHH 1 with their OPNs of 051 and 052, respectively. The ex-husband's sister appears as part of his SUBHH 2 with her OPN of 061. The friend has moved out and does not appear in either household.

Time 1

Household records

HHID=089012 ASUBHH=0

Respondent records

HHID=089012 PN=010 ASUBHH=0

HHID=089012 PN=020 ASUBHH=0

Household member/child records

HHID=089012 ASUBHH=0 OPN=101 (child)

HHID=089012 ASUBHH=0 OPN=102 (child)

HHID=089012 ASUBHH=0 OPN=051 (her brother)

HHID=089012 ASUBHH=0 OPN=052 (her brother)

HHID=089012 ASUBHH=0 OPN=061 (his sister)

HHID=089012 ASUBHH=0 OPN=080 (friend)

Time 2

Household records

HHID=089012 CSUBHH=1

HHID=089012 CSUBHH=2

Respondent records

HHID=089012 PN=010 CSUBHH=2  
HHID=089012 PN=020 CSUBHH=1  
Household member/child records  
HHID=089012 CSUBHH=1 OPN=101 (child)  
HHID=089012 CSUBHH=1 OPN=102 (child)  
HHID=089012 CSUBHH=1 OPN=051 (her brother)  
HHID=089012 CSUBHH=1 OPN=052 (her brother)  
HHID=089012 CSUBHH=2 OPN=101 (child)  
HHID=089012 CSUBHH=2 OPN=102 (child)  
HHID=089012 CSUBHH=2 OPN=061 (his sister)

### **13A9. Couple Divorces, One Respondent Remarries, Both Split-off Households Have New Members.**

Two respondents in a sample household are married at the time of the first cross-section. Each respondent is assigned a HHID of 090123 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

By the time of the second cross-section, the couple has divorced. She has moved in with her mother. He has married a woman with two children. At the second cross-section, both original respondents retain the HHID of 090123, but he is assigned a SUBHH of 1 and she is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. His new spouse and new stepchildren are assigned the HHID of 090123 and the SUBHH of 1. His new spouse is assigned a PN of 011. His new stepchildren are assigned OPNs of 151 and 152. Her mother is assigned the HHID of 090123 and the SUBHH of 2 and an OPN of 151.

Time 1

Household records  
HHID=090123 ASUBHH=0  
Respondent records  
HHID=090123 PN=010 ASUBHH=0  
HHID=090123 PN=020 ASUBHH=0

Time 2

Household records  
HHID=090123 CSUBHH=1  
HHID=090123 CSUBHH=2  
Respondent records  
HHID=090123 PN=010 CSUBHH=1  
HHID=090123 PN=011 CSUBHH=1 (new wife)  
HHID=090123 PN=020 CSUBHH=2  
Household member/child records  
HHID=090123 CSUBHH=1 OPN=151 (his stepchild)  
HHID=090123 CSUBHH=1 OPN=152 (his stepchild)  
HHID=090123 CSUBHH=2 OPN=151 (her mom)

### **13B. Cross-sectional Merging**

Many analyses require variables that are distributed in separate files. Sometimes you will need to obtain variables from files at different levels that contain different numbers of records. Before you can do your analysis work, the files will need to be merged in an appropriate manner. Prior to doing any data management you should ask yourself several questions.

- o What should be the level of the analysis file - will it have one record per respondent, one record per household or what?

- o What identification variables will be required to merge the various files that contain variables of interest for your analysis?
- o Will the merge be a one-to-one matching of records or a one-to-many matching?
- o What are the variables of interest? Predetermining which variables are needed for an analysis allows you to subset files to include only the necessary analytic and identification variables. Smaller files are more manageable.

We've provided sample setups for several types of merging. We hope that these will be of help to you as you think of your particular analytic needs.

### ***13B1. Respondent-level Files***

To create a respondent-level file with data from two or more respondent-level files, merge the respondent-level files using HHID and PN. This is a one-to-one match. Each input file contains 10,964 records. A respondent-level output file with 10,964 respondent records results.

### ***13B2. Household-level Files***

To create a household-level file with data from two or more household-level files, merge the household-level files using HHID and ESUBHH. This is a one-to-one match. Each input file contains 6,816 records. A household-level output file with 6,816 household records results.

### ***13B3. Respondent-level Files with Household-level Data***

To create a respondent-level file including data from a household-level file, merge the respondent-level file(s) and the household-level file(s) using HHID and ESUBHH. This is a one-to-many match (one household-to-many respondents). Household-level input files contain 6,816 records; respondent-level input files contain 10,964 records. A respondent-level output file with 10,964 respondent records results.

Since you are creating a respondent-level output file, be sure to keep PN, as well as HHID and ESUBHH, in the output file even though PN is not used for the merging.

### ***13B4. Household-level Files with Respondent-level Data***

To create a household-level file including variables from both respondents requires several steps.

First, separate the respondent-level records into two groups, one for the family respondent, the first person, and one for the non-family respondent, the second person.

We suggest you use EFAMR - 1996 WHETHER FAMILY RESPONDENT to create the two groups of respondents. The first group will contain 6,761 records (55 households did not have a family respondent); the second group will contain 4,203 records.

If you choose to use another variable to create two respondent groups, be sure each respondent group does not have more than one person from any one household (otherwise data will be lost).

Second, rename all variables for the second respondent group (except HHID and ESUBHH) to avoid overlap in the combined file.

Third, merge the two respondent files and the household-level file using HHID and ESUBHH.

Although the respondent groups contain fewer than 6,816 records, this is a one-to-one match. A household-level file with 6,816 household records results. The household-level output file includes household variables, a set of variables for the first respondent and a set of variables for the second respondent.

We recommend that you keep PN for the first and second respondent even though PN is not used for the merging and the output file is a household-level file.

### **13B5. Creating a Respondent-level File with Parent Data from a Household-level File**

Questions about parents of both the respondent and the respondent's spouse or partner were asked of the family respondent. This information is distributed at a household-level. You may wish to create a respondent-level file with information about the respondent's parents, whether the respondent was the family respondent or not. In order to do this, you will have to obtain the information about the parents from the household record and merge the variables to the proper respondent record.

In the example below, the file "famr" will have 6,761 observations one for each family respondent. The file "nfamr" will have 4,203 observations one for each non-family respondent. The combined file "resp" will have 10,964 observations one for each respondent.

### **13B6. SAS Examples**

#### **13B6a. SAS Example of Respondent to Respondent Merge**

```
proc sort
    data=in.h96a_r
    out=h96a_r(keep=hhid pn e638 e639);
    by hhid pn;
run;
proc sort
    data=in.h96b_r
    out=h96b_r(keep=hhid pn e769 e772);
    by hhid pn;
run;
proc sort
    data=in.h96c_r
    out=h96c_r(keep=hhid pn e1161 e1162);
    by hhid pn;
run;
data resp;
    merge h96a_r h96b_r h96c_r;
    by hhid pn;
run;
```

#### **13B6b. SAS Example of Household to Household Merge**

```
proc sort
    data=in.h96f_h
    out=h96f_h(keep=hhid esubhh e2224 e2225);
    by hhid esubhh;
run;
```

```

proc sort
    data=in.h96j_h
    out=h96j_h(keep=hhid esubhh e3875 e3877);
    by hhid esubhh;
run;
data hhold;
    merge h96f_h h96j_h;
    by hhid esubhh;
run;

```

### 13B6c. SAS Example of Household to Respondent Merge

```

proc sort
    data=in.h96a_r
    out=h96a_r(keep=hhid esubhh pn e638 e639);
    by hhid esubhh;
run;
proc sort
    data=in.h96j_h
    out=h96j_h(keep=hhid esubhh e3875 e3877);
    by hhid esubhh;
run;
data resp;
    merge h96a_r h96j_h;
    by hhid esubhh;
run;

```

### 13B6d. SAS Example of Respondent to Household Merge

```

* file of family respondents;
proc sort
    data=in.h96a_r(keep=hhid esubhh efamr pn e638 e639
    rename=(efamr=efamra pn=pna e638=e638a e639=e639a)
    where=(efamra in(1 2)))
    out=h96a1;
    by hhid esubhh;
run;
* file of non-family respondents;
proc sort
    data=in.h96a_r(keep=hhid esubhh efamr pn e638 e639
    rename=(efamr=efamrb pn=pnb e638=e638b e639=e639b)
    where=(efamrb not in(1 2)))
    out=h96a2;
    by hhid esubhh;
run;
* variables from a household-level file;
proc sort
    data=in.h96j_h(keep=hhid esubhh e3875 e3877)
    out=h96j;
    by hhid esubhh;
run;
* household-level file with information from two respondents;
data hhold;
    merge
    h96a1 h96a2 h96j;
    by hhid esubhh;
run;

```

### 13B6e. SAS Example of Respondent File with Parent Information from a Household-level File

```
*      select parent variables from household file for family r;
data famr;
  set in.h96d_h;
  *      keep id variables;
  keep hhid pn;
  *      keep parent variables;
  e1557_1 e1558_1;
  *      assign person number for family r;
  pn=epn_fam;
  attrib pn label='PERSON NUMBER' format=$char3.;
  *      output records for households with family R;
  if pn ne '';

run;
*      select parent variables from household file for non-family r;
data nfamr;
  set in.h96d_h;
  *      keep id variables;
  keep hhid pn;
  *      keep parent variables;
  e1557_1 e1558_1;
  *      copy in-law variables in to output variable location;
  e1557_1=e1557_2;
  e1558_1=e1558_2;
  *      assign person number for family r;
  pn=epn_nfam;
  attrib pn label='PERSON NUMBER' format=$char3.;
  *      output records for households with non-family R;
  if pn ne '';

run;
*      concatenate files;
data resp;
  set famr nfamr;

run;
```

### 13B7. Stata Examples

#### 13B7a. Stata Example of Respondent to Respondent Merge

```
set prefix "hrs1996"
* This is an optional statement
* Make sure all the data files are in directory "hrs1996"

use HHID PN E769 E772 using h96b_r
sort HHID PN
save tmp1, replace

use HHID PN E1161 E1162 using h96c_r
sort HHID PN
save tmp2, replace

use HHID PN E638 E639 using h96a_r
sort HHID PN
merge HHID PN using tmp1
drop _m
sort HHID PN
merge HHID PN using tmp2
```



```
drop _m
sort HHID PN
save resp, replace
```

#### 13B7b. Stata Example of Household to Household Merge

```
use HHID ESUBHH E2224 E2225 using h96f_h
sort HHID ESUBHH
save tmp, replace
```

```
use HHID ESUBHH E3875 E3877 using h96j_h
sort HHID ESUBHH
merge HHID ESUBHH using tmp
drop _m
sort HHID ESUBHH
save hhold, replace
```

#### 13B7c. Stata Example of Household to Respondent Merge

```
use HHID PN ESUBHH E769 E772 using h96b_r
sort HHID PN
save tmp1, replace
```

```
use HHID ESUBHH E3875 E3877 using h96j_h
sort HHID ESUBHH
save tmp2, replace
```

```
use HHID PN ESUBHH E638 E639 using h96a_r
sort HHID PN
merge HHID PN using tmp1
tab _m
drop _m
sort HHID ESUBHH
merge HHID ESUBHH using tmp2
tab _m
drop _m
sort HHID PN
save resp, replace
```

#### 13B7d. Stata Example of Respondent to Household Merge

```
use HHID ESUBHH E3875 E3877 using h96j_h
sort HHID ESUBHH
save tmp1, replace
```

```
use HHID ESUBHH EFAMR PN E638 E639 using h96a_r
keep if EFAMR==1|EFAMR==2
rename EFAMR EFAMRa
rename PN PNa
rename E638 E638a
rename E639 E639a
sort HHID ESUBHH
save tmp2, replace
```

```
use HHID ESUBHH EFAMR PN E638 E639 using h96a_r
keep if EFAMR~=1 & EFAMR~=2
rename EFAMR EFAMRb
rename PN PNb
rename E638 E638b
```

```

rename E639 E639b
sort HHID ESUBHH
merge HHID ESUBHH using tmp2
tab _m
drop _m
sort HHID ESUBHH
merge HHID ESUBHH using tmp1
tab _m
drop _m
sort HHID ESUBHH
save hhold, replace

```

### 13B7e. Stata Example of Respondent File with Parent Information from a Household-level File

```

* select parent variables from household file for family r
use h96d_h, clear

* assign person number for family r
gen str3 PN=EPN_FAM

* keep ID variables and parent variables of interest
keep HHID PN E1557_1 E1558_1

* out records for family respondents
keep if PN~=""
save fam1, replace

* select parent variables from household file for non-family r
use h96d_h, clear
drop E1557_1 E1558_1

* assign person number for non-family r
gen str3 PN=EPN_FAM

* copy in-law variables in to output variable location
gen E1557_1=E1557_2
gen E1558_1=E1558_2

* keep ID variables and parent variables of interest
keep HHID PN E1557_1 E1558_1

* out records for non-family respondents
keep if PN~=""
save fam2, replace

* concatenate files
use fam1, clear
append using fam2
save resp, replace

```

### 13B8. SPSS Examples

#### 13B8a. SPSS Example of Respondent to Respondent Merge

```

/*-----
/* sort each input dataset by HHID PN
/* subset variables

```

```

GET FILE 'c:\hrs1996\spss\h96a_r.sav'
  /keep=hhid pn e638 e639.
SORT CASES BY hhid pn.
save outfile='c:\temp\h96a_r.sav'.

GET FILE 'c:\hrs1996\spss\h96b_r.sav'
  /keep=hhid pn e769 e772.
SORT CASES BY hhid pn.
save outfile='c:\temp\h96b_r.sav'.

GET FILE 'c:\hrs1996\spss\h96c_r.sav'
  /keep=hhid pn e1161 e1162.
SORT CASES BY hhid pn.
save outfile='c:\temp\h96c_r.sav'.

/*-----
/*      merge three datasets by HHID PN to create a respondent file
/*      this is a one-to-one merge

MATCH FILES
  /FILE='c:\temp\h96a_r.sav'
  /FILE='c:\temp\h96b_r.sav'
  /FILE='c:\temp\h96c_r.sav'
  /BY hhid pn.
EXECUTE.

save outfile='c:\hrs1996\spss\resp.sav'.

```

### 13B8b. SPSS Example of Household to Household Merge

```

/*-----
/*      sort each input dataset by HHID ESUBHH
/*      subset variables

GET FILE 'c:\hrs1996\spss\h96f_h.sav'
  /keep=hhid esubhh e2224 e2225.
SORT CASES BY hhid esubhh.
save outfile='c:\temp\h96f_h.sav'.

GET FILE 'c:\hrs1996\spss\h96j_h.sav'
  /keep=hhid esubhh e3875 e3877.
SORT CASES BY hhid esubhh.
save outfile='c:\temp\h96j_h.sav'.

/*-----
/*      merge two datasets by HHID ESUBHH to create a household file
/*      this is a one-to-one merge

MATCH FILES
  /FILE='c:\temp\h96f_h.sav'
  /FILE='c:\temp\h96j_h.sav'
  /BY hhid esubhh.
EXECUTE.

save outfile='c:\hrs1996\spss\hhold.sav'.

```

### 13B8c. SPSS Example of Household to Respondent Merge

```
/*-----  
/*      sort each input dataset by HHID ESUBHH  
/*      subset variables  
  
GET FILE 'c:\hrs1996\spss\h96a_r.sav'  
      /keep=hhid esubhh pn e638 e639.  
SORT CASES BY hhid esubhh.  
save outfile='c:\temp\h96a_r.sav'.  
  
GET FILE 'c:\hrs1996\spss\h96j_h.sav'  
      /keep=hhid esubhh e3875 e3877.  
SORT CASES BY hhid esubhh.  
save outfile='c:\temp\h96j_h.sav'.  
  
/*-----  
/*      merge two datasets by HHID ESUBHH to create a respondent file  
/*      this is a one-to-many merge  
  
MATCH FILES  
      /FILE='c:\temp\h96a_r.sav'  
      /TABLE='c:\temp\h96j_h.sav'  
      /BY hhid esubhh.  
EXECUTE.  
  
save outfile='c:\hrs1996\spss\resp.sav'.
```

### 13B8d. SPSS Example of Respondent to Household Merge

```
/*-----  
/*      file of family respondents  
  
GET FILE 'c:\hrs1996\spss\h96a_r.sav'  
      /keep=hhid esubhh efamr pn e638 e639.  
USE ALL.  
SELECT IF(efamr = 1 or efamr = 2 or efamr = 3 or efamr=4).  
sort cases by hhid esubhh.  
rename variables efamr=efamra.  
rename variables pn=pna.  
rename variables e638=e638a.  
rename variables e639=e639a.  
EXECUTE .  
  
save outfile='c:\temp\h96a1.sav'.  
  
/*-----  
/*      file of non-family respondents  
  
GET FILE 'c:\hrs1996\spss\h96a_r.sav'  
      /keep=hhid esubhh efamr pn e638 e639.  
sort cases by hhid esubhh.  
USE ALL.  
SELECT IF(efamr <> 1 and efamr <> 2 and efamr <> 3 and efamr<>4).  
rename variables efamr=efamrb.  
rename variables pn=pnb.  
rename variables e638=e638b.  
rename variables e639=e639b.
```

```

EXECUTE .
save outfile='c:\temp\h96a2.sav'.

/*-----
/*    variables from a household-level file

GET FILE 'c:\hrs1996\spss\h96j_h.sav'
      /keep=hhid esubhh e3875 e3877.
SORT CASES BY hhid esubhh.
save outfile='c:\temp\h96j.sav'.

/*-----
/*    household-level file with information from two respondents

MATCH FILES
      /FILE='c:\temp\h96a1.sav'
      /FILE='c:\temp\h96a2.sav'
      /FILE="c:\temp\h96j.sav"
      /BY hhid esubhh.
EXECUTE.

```

```
save outfile='c:\hrs1996\spss\hhold.sav'.
```

### 13B8e. SPSS Example of Respondent File with Parent Information from a Household-level File

```

/*-----
/*    select parent variables from household file for family r

/*    keep id variables and parent variables of interest
GET FILE 'c:\hrs1996\spss\h96d_h.sav'
      /keep=hhid epn_fam e1557_1 e1558_1.

/*    assign person number for family r
STRING pn (A3).
COMPUTE pn = epn_fam .
VARIABLE LABELS pn 'Person Number' .

USE ALL.
/*    output records for households with family R
SELECT IF(pn <>'').
EXECUTE.
save outfile='c:\temp\famr.sav' /drop= epn_fam.

/*-----
/*    select parent variables from household file for non-family r

/*    keep id variables and parent variables of interest
GET FILE 'c:\hrs1996\spss\h96d_h.sav'
      /keep=hhid epn_nfam e1557_2 e1558_2.

/*    assign person number for family r
STRING pn (A3).
COMPUTE pn = epn_nfam .
VARIABLE LABELS pn 'Person Number' .

/*    copy in-law variables in to output variable location
rename variables e1557_2=e1557_1.

```

```
rename variables e1558_2=e1558_1.
USE ALL.

/*      output records for households with non-family R
SELECT IF(pn <>'').
EXECUTE .

save outfile='c:\temp\nfamr.sav' /drop= epn_nfam.

/*-----
/*      concatenate files
ADD FILES
      /FILE='c:\temp\nfamr.sav'
      /FILE='c:\temp\famr.sav'.
EXECUTE.

save outfile='c:\hrs1996\spss\resp.sav'.
```