

HEALTH AND RETIREMENT STUDY 1998
Final Core Data Release, Version 2
May 2003
(Revised October 2003)
Data Description and Usage

TABLE OF CONTENTS

1.	OVERVIEW.....	5
1A.	WHAT'S NEW IN VERSION 2	5
1B.	THE SAMPLE INTERVIEWED IN 1998	5
1C.	1998 QUESTIONNAIRE SECTIONS	5
1D.	LEVELS OF FILES.....	6
2.	FILE NAMING CONVENTIONS	7
3.	DATA FILES.....	8
4.	IDENTIFICATION VARIABLES	9
4A.	PRIMARY IDENTIFICATION VARIABLES.....	9
4B.	PRIMARY IDENTIFICATION VARIABLES FOR DATASETS AT EACH OF THE SEVEN LEVELS	10
4C.	SECONDARY IDENTIFICATION VARIABLES.....	11
4D.	DATASETS INCLUDING FSUBHH AND OPN AS PRIMARY IDENTIFIERS	11
5.	DISTRIBUTION FILES AND DIRECTORY STRUCTURE	12
5A.	DISTRIBUTION FILES.....	12
5B.	DIRECTORY STRUCTURE	12
6.	PROGRAM STATEMENTS	13
6A.	USING THE FILES WITH SAS	13
6B.	USING THE FILES WITH STATA.....	13
6C.	USING THE FILES WITH SPSS.....	13
7.	DOCUMENTATION	14
7A.	CODEBOOK.....	14
7B.	OTHER TYPES OF DOCUMENTATION	21
8.	ADDITIONAL NOTES	21
8A.	BASIC DEMOGRAPHIC VARIABLES	21
8B.	SELECTING THE CORRECT VARIABLE FOR THE RESPONDENT FROM PRELOAD INFORMATION.....	22
8C.	HOUSEHOLDS WITH NO OR TWO COVERSHEET OR FAMILY OR FINANCIAL RESPONDENTS	24
8D.	UNFOLDING BRACKET VARIABLES AND IMPUTATIONS	25
8E.	CHANGES IN DATA BECAUSE OF RESPONDENT COMMENTS	25
8F.	MODULES FOR THE 1998 DATA COLLECTION.....	26
8G.	TWO VERSIONS OF MASKED OCCUPATION CODE.....	26
8H.	VARIABLES NOT INCLUDED PUBLICLY RELEASED	27
8I.	DETAILED COMPARISON OF VARIABLES IN VERSION 1 AND VERSION 2	27
9.	LONGITUDINAL ISSUES.....	28
9A.	THE SAMPLE DESIGN	28
9B.	LINKING RESPONDENTS ACROSS TIME.....	28
9C.	LINKING OTHER PERSONS ACROSS TIME.....	28
9D.	SUMMARY OF CHANGES FROM PRIOR WAVES NOT NEW FOR 1998	29
10.	OTHER DATA FILES	31
10A.	TRACKER FILE.....	31

10B.	IMPUTATION FILE	32
11.	OBTAINING THE DATA	32
11A.	DOWNLOADING THE DATA	32
11B.	REGISTRATION AND CONDITIONS OF USE.....	32
12.	IF YOU NEED TO KNOW MORE.....	33
12A.	HRS INTERNET SITE	33
12B.	CONTACT INFORMATION.....	33
13.	EXAMPLES OF MERGING	34
13A.	RESPONDENT-LEVEL FILES	34
13B.	HOUSEHOLD-LEVEL FILES	34
13C.	RESPONDENT-LEVEL FILES WITH HOUSEHOLD-LEVEL DATA	34
13D.	HOUSEHOLD-LEVEL FILES WITH RESPONDENT-LEVEL DATA	34
13E.	CREATING A RESPONDENT-LEVEL FILE WITH PARENT DATA FROM A HOUSEHOLD-LEVEL FILE.....	35
13F.	SAS EXAMPLES	35
13G.	STATA EXAMPLES.....	37
13H.	SPSS EXAMPLES.....	39
14.	EXAMPLES OF SUB-HOUSEHOLD AND RESPONDENT PERSON NUMBER AND OTHER PERSON NUMBER ASSIGNMENTS.....	43
14A.	MARRIED COUPLE STAYS MARRIED.....	43
14B.	COUPLE DIVORCES.....	44
14C.	ONE OR BOTH RESPONDENTS DIE.....	44
14D.	SINGLE RESPONDENT MARRIES.	44
14E.	COUPLE DIVORCES, ONE RESPONDENT REMARRIES AND DIVORCES.	44
14F.	COUPLE DIVORCES AND MARRIES AGAIN.....	45
14G.	MARRIED COUPLE WITH CHILDREN AND SIBLINGS.	45
14H.	COUPLE WITH CHILDREN AND SIBLINGS DIVORCES.....	45
14I.	COUPLE DIVORCES, ONE RESPONDENT REMARRIES, BOTH SPLIT-OFF HOUSEHOLDS HAVE NEW MEMBERS.....	45

1. Overview

The HRS 1998 Final Core 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. What's New in Version 2

A number of minor changes have been incorporated into the version 2 data release.

- o A few of newly coded open-ended variables are included in the version 2 release. Some variables previously included in the respondent-level coversheet file are now included in a household-level coversheet file instead.¹ A number of other variables were added or corrected. See [Detailed Comparison of Variables in Version 1 and Version 2](#) for more information.
- o The version 2 codebook includes [assignment statements](#) and additional user notes that document how values were assigned for variables created by the data collection program.
- o This document includes [SPSS Examples](#) in the Examples of Merging section.

1B. The Sample Interviewed in 1998

The 1998 data collection combined for the first time the AHEAD and the HRS sub-samples and added two additional sub-samples. The AHEAD sub-sample was interviewed previously in 1993 and 1995. The HRS sub-sample was interviewed previously in 1992, 1994 and 1996. In 1998, and every two years thereafter, the HRS, AHEAD and the two new sub-samples - War Babies (WB), and Children of the Depression (CODA) - were interviewed. For more details about the sub-samples, see the section [The Sample Design](#) below. The data collection period for the 1998 interview was February 1998 through March of 1999.

1C. 1998 Questionnaire Sections

The 1998 HRS data collection instrument or questionnaire contained the following sections.

Section PR Preload
Section CS Cover Screen

¹ The version 2, May 2003, release includes three coversheet files - one at the respondent-level, one at the household-member-child-level, and one at the household-level. The version 1, December 2001, release included two coversheet files - one at the respondent-level and one at the household-member-child-level.

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

See the section [Modules for the 1998 Data Collection](#) below for details about the 1998 modules.

1D. Levels of Files

In the 1998 data collection instrument, most questions were asked of all respondents. For two respondent households, some questions were asked just of a designated financial respondent on behalf of the entire household. Other questions were just asked of a designated family respondent on behalf of the entire household. And some questions were just asked of the first respondent interviewed, the coversheet respondent, on behalf of the entire household.

In addition to the familiar household-level and respondent-level files, the HRS 1998 Final Core Data Release, version 2, contains files at five other levels -- household-member-and-child, sibling, helper, transfer-to-child and transfer-from-child.

1D1. Household-Level Files

Household-level files contain questions that were asked about the household of a designated coversheet, financial or family respondent. A coversheet respondent answered family questions (section CS) on behalf of the entire household. Similarly, a family respondent answered family questions (section D) on behalf of the entire household and a financial respondent answered household-level financial questions (sections F, J, and N) on behalf of the entire household. The household-level files contain one record for each household in which at least one interview was obtained in 1998.

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 give an interview). The files one record for each respondent for whom an interview was obtained in 1998.

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 one record 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 of the respondent's spouse/partner (sections PR and D). If the designated family respondent did not provide an interview, information from the non-family respondent was used for PR sibling file.

Detailed questions about siblings were only asked if the respondent's mother or father was alive this wave or was alive the previous wave. Similarly, detailed questions about siblings of the spouse/partner were asked only if the spouse's/partner's mother or father was alive this wave or was alive the previous wave.²

The files contain one record for each sibling or spouse's/partner's sibling about whom detailed questions were asked.

1D5. Helper-Level File

This file contains information provided by each respondent about help received from a helper 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 one record for help received by the respondent from each helper or, if the helper is a married or partnered child, the helping couple.

1D6. Transfer-to-Child-Level File

This file contains information provided by the family respondent about transfers of money to a child, children, grandchild or grandchildren. The file contains one record for each transfer.

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

This file contains information provided by the family respondent about transfers of money from to a child, children, grandchild or grandchildren. The file contains one record for each transfer.

2. File Naming Conventions

In the version 2 release, files are named beginning with "H98" for HRS 1998. Next a letter or two designating the questionnaire section. Next a separator, "_" and then 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-level, and
- FC for transfer-from-child-level.

Putting it all together, files named H98A_R include variables from section A at the respondent-level.

² Detailed questions were also asked about siblings of a spouse/partner who died during the course of the study, if the respondent did not have a new spouse/partner, and if the deceased spouse's/partner's mother or father was alive this wave or was alive the previous wave.

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 provided for each of 29 parts of the 1998 HRS Final Core Data Release, version 2, listed below. For example, H98A_R.DA contains data, H98A_R.SAS contains corresponding SAS program statements, H98A_R.SPS contains corresponding SPSS program statements, H98A_R.DO contains corresponding Stata DO statements, H98A_R.DCT contains corresponding Stata dictionary statements, and H98A_R.TXT contains the ASCII codebook.

3. Data Files

The HRS 1998 Final Core Data Release, version 2, is distributed in 29 parts. The parts are listed below along with the number of cases (N) and number of variables (NV). The files are sorted by the designated primary identifiers.

HRS 1998 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.

Household-level-files

Sorted by: HHID FSUBHH
N=14395

H98CS_H	NV=67
H98D_H	NV=523
H98F_H	NV=175
H98J_H	NV=596
H98N_H	NV=112

Respondent-level-files

Sorted by: HHID PN
N=21384

H98PR_R	NV=257
H98CS_R	NV=86
H98A_R	NV=96
H98B_R	NV=190
H98C_R	NV=95
H98D_R	NV=16
H98E_R	NV=295
H98G_R	NV=1480
H98H_R	NV=53
H98J_R	NV=54
H98M_R	NV=191
H98N_R	NV=82
H98PC_R	NV=74
H98R_R	NV=184
H98T_R	NV=49

Household-member-and-child-level-files

Sorted by: HHID FSUBHH OPN
N=49013

H98PR_MC NV=35
H98CS_MC NV=18
H98D_MC NV=48
H98J_MC NV=28

Sibling-level-files

Sorted by: HHID FSUBHH OPN
N=21409

H98PR_SB NV=22
H98D_SB NV=35

Transfer-to-child-level-file

Sorted by: HHID FSUBHH OPN FTRANNUM
N=6176

H98D_TC NV=27

Transfer-from-child-level-file

Sorted by: HHID FSUBHH OPN FTRANNUM
N=797

H98D_FC NV=27

Helper-level-file

Sorted by: HHID PN OPN
N=3455

H98E_HP NV=49

4. Identification Variables

Identification variables for HRS 1998 are stored in character format. For further information about merging 1998 data, see [Examples of Merging](#) below.

4A. Primary Identification Variables.

Five variables, HHID, FSUBHH, PN, OPN, and FTRANNUM are used in various combinations to uniquely identify the five different level datasets that comprise this data release.

4A1. HHID – Household Identifier.

At the initial wave of data collection (in 1992 for the HRS sub-sample, in 1993 the AHEAD sub-sample, or 1998 for the WB and CODA sub-samples) each sample household was assigned a Household Identifier. 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. FSUBHH – HRS 1998 Sub-household Identifier.

In combination with HHID, FSUBHH uniquely identifies a household at the time of the 1998 data collection. Sub-household identifiers can be different at each

wave. FSUBHH has one-digit. For more information, see [Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments](#) below.

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 FSUBHH, OPN uniquely identifies another person - household member, child, or sibling at the time of the 1998 data collection. In combination with HHID and PN, OPN uniquely identifies a help given at the time of the 1998 data collection. OPN has three-digits. See [Datasets Including FSUBHH and OPN as Primary Identifiers](#) for additional information.

4A5. *FTRANNUM – 1998 TYPE OF TRANSFER*

In combination with HHID and FSUBHH, OPN uniquely identifies a transfer to or from a child, children, grandchild or grandchildren. FTRANNUM has one digit.

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

Two identifiers uniquely identify records in the

- o respondent-level datasets:

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

Three identifiers uniquely identify records in the

- o helper-level datasets:

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

Two identifiers uniquely identify records in the

- o household-level datasets:

- 1) HHID HOUSEHOLD IDENTIFIER and
- 2) FSUBHH 1998 SUB-HOUSEHOLD IDENTIFIER.

Three identifiers uniquely identify records in the

- o household-member-and-child-level, and
- o sibling-level datasets:

- 1) HHID HOUSEHOLD IDENTIFIER,
- 2) FSUBHH 1998 SUB-HOUSEHOLD IDENTIFIER, and
- 3) OPN OTHER PERSON NUMBER.

Four identifiers uniquely identify records in the

- o transfer-to-child-level, and
- o transfer-from-child-level datasets:

- 1) HHID HOUSEHOLD IDENTIFIER,
- 2) FSUBHH 1998 SUB-HOUSEHOLD IDENTIFIER,

- 3) OPN OTHER PERSON NUMBER, and
- 4) FTRANNUM1998 TRANSFER NUMBER

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

- o respondent-level and
- o helper-level datasets:

FSUBHH	1998	SUB-HOUSEHOLD IDENTIFIER
ESUBHH	1996	SUB-HOUSEHOLD IDENTIFIER
DSUBHH	1995	SUB-HOUSEHOLD IDENTIFIER
FPN_SP	1998	SPOUSE/PARTNER PERSON NUMBER
FCSR	1998	WHETHER COVERSHEET RESPONDENT
FFAMR	1998	WHETHER FAMILY RESPONDENT
FFINR	1998	WHETHER FINANCIAL RESPONDENT
FQNR	1998	QUESTIONNAIRE

Secondary Identification Variables for

- o household-level,
- o household-member-or-child-level,
- o sibling-level,
- o transfer-from-child-level, and
- o transfer-to-child-level datasets:

ESUBHH	1996	SUB-HOUSEHOLD IDENTIFIER
DSUBHH	1995	SUB-HOUSEHOLD IDENTIFIER
FPN_CS	1998	COVERSHEET RESP PERSON NUMBER
FPN_FAM	1998	FAMILY RESP PERSON NUMBER
FPN_FIN	1998	FINANCIAL RESP PERSON NUMBER
FPN_NCS	1998	NON-COVERSHEET RESP PERSON NUMBER
FPN_NFAM	1998	NON-FAMILY RESP PERSON NUMBER
FPN_NFIN	1998	NON-FINANCIAL RESP PERSON NUMBER
FQNR_CS	1998	QUESTIONNAIRE - COVERSHEET RESP
FQNR_FAM	1998	QUESTIONNAIRE - FAMILY RESP
FQNR_FIN	1998	QUESTIONNAIRE - FINANCIAL RESP

4D. Datasets Including FSUBHH and OPN as Primary Identifiers

When working with datasets including FSUBHH and OPN as primary identifiers -- household-member-and-child-level, sibling-level, transfer-from-child-level, or transfer-to-child-level datasets - it is essential to use FSUBHH 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 in [Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments](#) where the couple with children divorces, the record where HHID is 89012, FSUBHH is 1, and OPN is 101 is the report about the child from the mother, and the record where HHID is 89012, FSUBHH 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 where the couple divorces and both split-off households have new members, the record where HHID is 90123, FSUBHH is 1, and OPN is 151 is for the ex-wife's mother, while the record where HHID is 90123, FSUBHH is 2, and OPN is 151 is for the ex-husband's new stepchild.

See also the section [Linking Other Persons across Time](#) below.

5. Distribution Files and Directory Structure

5A. Distribution Files

The files are packaged for download from our Web site in two different ways - as one big .ZIP file that contains seven smaller .ZIP files and a .PDF version of this document or the seven smaller .ZIP files available individually for separate download. You'll typically want to download at least the data files, one set of program statement files, and the codebook files.

The combined file is H98CORE.ZIP. The individual files for separate download are:

- Data files
 - H98DA.ZIP containing data files.
- Program statement files
 - H98SAS.ZIP containing SAS data descriptors.
 - H98SPS.ZIP containing SPSS data descriptors.
 - H98STA.ZIP containing Stata data descriptors.
- Documentation files
 - H98CB.ZIP containing codebook.
 - H98QN.ZIP containing questionnaire.
 - H98SPC.ZIP containing Surveycraft programming specifications.

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 distribution files.

Directory	Contents
c:\HRS1998	Files downloaded from Web site
c:\hrs1998\codebook	Unzipped files from h98cb.zip
c:\hrs1998\data	Unzipped files from h98da.zip
c:\hrs1998\qnaire	Unzipped files from h98qn.zip
c:\hrs1998\sas	Unzipped files from h98sas.zip
c:\hrs1998\spss	Unzipped files from h98sps.zip
c:\hrs1998\stata	Unzipped files from h98sta.zip
c:\hrs1998\scraft	Unzipped files from h98prn.zip

Decompress the selected ZIP files into the appropriate subdirectories. You will need about 205 MB of free space on your storage device to store the 29 .DA files.

6. Program Statements

Each data file comes with associated SAS, Stata or SPSS 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 H98A_R.SAS go with the H98A_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:\hrs1998\SAS" and the data file is located in "c:\hrs1998\DATA", you can run the file as is. A SAS system file (*.SD2 or *.SAS7BDAT) will be saved to directory "c:\hrs1998\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 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:\hrs1998\STATA" and the data file is located in "c:\hrs1998\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.

6C. 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:\hrs1998\SPSS" and the data file is located in "c:\hrs1998\DATA", you can run the file as is. An SPSS system file (*.SAV) will be saved to directory "c:\hrs1998\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.

7. Documentation

There are several types of documentation available for use with the HRS 1998 Final Core Data Release. In addition to this document, they include a codebook, programming specifications, and a box-and-arrow questionnaire.

7A. Codebook

The HRS 1998 Codebook is provided as a series of 29 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.

```
=====
F1097      B1. RATE HEALTH
Section: B           Level: Respondent           CAI Reference: Q1097
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?
.....
2632              1. EXCELLENT
5467              2. VERY GOOD
6542              3. GOOD
4400              4. FAIR
2337              5. POOR
                  7. Other
                   6           8. DK (don't know); NA (not ascertained)
                   9           9. RF (refused)
Blank. INAP (Inapplicable): [Q456:CS CONTINUE] IS
(5); [Q497:CS2b] IS (A) OR [Q542:CS15C2] IS
(A) OR [Q518:CS11a] IS (A)
=====
```

ASSIGNMENT STATEMENTS

- * If {When Q160 is (Q453), [Q183:VIRGIN PREV WAVE FINAN-NONFIN] IS (1) AND [Q182:VIRGIN PREV WAVE FAMILY-NONFAM] IS (5)} then [F544.VIRGIN R FAM/FIN] = 1
- * If {When Q160 is (Q453), [Q183:VIRGIN PREV WAVE FINAN-NONFIN] IS (5) AND [Q182:VIRGIN PREV WAVE FAMILY-NONFAM] IS (1)} then [F544.VIRGIN R FAM/FIN] = 2
- * If {When Q160 is (Q453), [Q183:VIRGIN PREV WAVE FINAN-NONFIN] IS (1) AND [Q182:VIRGIN PREV WAVE FAMILY-NONFAM] IS (1)} then [F544.VIRGIN R FAM/FIN] = 3
- * If {When Q160 is (Q453), [Q183:VIRGIN PREV WAVE FINAN-NONFIN] IS (5) AND [Q182:VIRGIN PREV WAVE FAMILY-NONFAM] IS (5)} then [F544.VIRGIN R FAM/FIN] = 4

```
-----
F544      VIRGIN R FAM/FIN
Section: CS           Level: Respondent           CAI Reference: Q544
Type: Numeric       Width: 1                   Decimals: 0
```

User Note: VIRGIN R FAM/FIN is assigned a value of 0.

```
.....
4961          0. NOT INTERVIEWED
3177          1. FINANCIAL R
3220          2. FAMILY R
```

7621 3. FINANCIAL & FAMILY R
 2405 4. NON-FINANCIAL & NON-FAMILY R
 Blank. INAP (Inapplicable): [Q456:CS CONTINUE] IS (5);
 [Q497:CS2b] IS (A) OR [Q542:CS15C2] IS (A) OR
 [Q518:CS11a] IS (A)

=====

F2760	F6.HOME VALUE \$		
	Section: F	Level: Household	CAI Reference: Q2760
	Type: Numeric	Width: 8	Decimals: 0

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

DO NOT PROBE DK

0 NOTHING

AMOUNT:

.....

8	0. Actual value
23	1-4999. Actual value (verified)
8387	5000-500000. Actual value
104	500001-99999996. Actual value (verified)
	99999997. Other
1224	99999998. DK (don't know); NA (not ascertained)
88	99999999. RF (refused)
4560	Blank. INAP (Inapplicable): [Q456:CS CONTINUE] IS (5); [Q497:CS2b] IS (A) OR [Q542:CS15C2] IS (A) OR [Q518:CS11a] IS (A); [Q689:R FIN-NON V106] IS (5); [Q517:CS11] IS (1) AND [Q721:CS26] IS (NE 1); [Q2743:F3] IS (2); [Q2743:F3] IS (7 OR DK OR RF); [Q2746:F4] IS (3); [Q2746:F4] IS (7 OR DK OR RF); [Q2751:F5] IS (4); [Q2751:F5] IS (7 OR DK OR RF); [Q2753:F5b] IS (NE DK AND NE RF); [Q2754:F5c] IS (3 OR DK OR RF); [Q2755:F5d] IS (NE 5); [Q2756:F5e] IS (A); [Q2757:F5f] IS (A); partial interview; no financial respondent

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 and User Notes with Assignments

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, they 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.

User notes of the form "User note: CS15Y11.VIRGIN R FAM/FIN is assigned a value of 0." denote the assignment of a value to a variable if the variable did not receive a value from a previous assignment statements, if any.

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 HRS 1998 cross-section begin with an "F", i.e., " F769". 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.

7A1b1. 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.

7A1b2. 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-level files,
- "FromChild" for transfer-from-child-level files, and
- "Helper" for helper-level files.

7A1b3. CAI Reference

The CAI (Computer Assisted Interviewing) Reference begins with the letter "Q". For data collected as part of the interview and not masked, 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 6697. The 10000 and 30000 series are used for masked variables, and derived variables with alternative formats. The 9000 series is used as the CAI Reference for identification variables. The 20000 series is used for data release version variables.

7A1b4. Type

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

7A1b5. Width

Variable width in the ASCII data file.

7A1b6. Decimals

The number of decimals. Usually zero.

7A1c. 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. Question text in upper and lower case was that read to the respondent. Interviewer instructions that appeared on the CATI screen but were not read to the respondent and rules that determined what text was read to the respondent are in upper case. Text fills, e.g., answers to previous questions that were part of the question text are designated with a Q-number and question reference or brief description, e.g., "since Q218-PREV WAVE IW MONTH".

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., "1992-1998"; 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)".

7A2c1. 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.

When used in the codebook, typically in the INAP text, but also in question text that includes rules for alternative wording, the (A) symbol means "answered". For example, "[Q3140:G5c] IS (A)" can be read as "if Q3140 was answered".

The text (NE "|") means "not blank". For example, "[Q3664:GG13] IS (NE "|)" can be read as "if Q3664 is not blank".

In the INAP text, 0 means a value of zero or blank or not answered. For instance, the phrase "[Q4589:H7] IS (00)" would include both a value of 0 and a value of blank for F4589, the variable with the CAI reference Q4589.

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, occasional, minor changes made to variable labels in processing.

7A3. Special Considerations

Several special codebook conventions and types of variables are described below.

7A3a. 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., F1022M1, F1022M2, F1022M3.

Variables from questions that were in a loop have an "_" in their name and a suffix that designates the loop, e.g., F1776_1, F1776_2, F1776_3.

Null multiple mention variables and variables from null loops are not included in the data. User notes indicating the number of mentions or loops possible and the number with non-null data are provided, e.g., "User note: Up to twelve responses were allowed for this question, the actual maximum number of responses was six."

Variables from questions in loops that allowed multiple mentions end with a three-digit number indicating the loop and the mention. For example, fifty variables would result for a question that was in a loop of five and allowed up to ten mentions. The resulting variables would be, for example, F2042001 to F2042010 for the ten mentions from the first loop, F2042011 to F2042020 for the ten mentions from the second loop and so forth, with F2042041 to F2042050 for the ten mentions from the fifth loop.

A user note is provided indicating the maximum and the actual number of loops and mentions, e.g., "User note: Up to five loops with ten mentions were allowed for this question. The maximum number of loops used was two and the maximum number of responses was six."

7A3b. Masked Variables

Because we are dedicated to protecting the anonymity of our respondents and to protecting the confidentiality of the information they provide, a number of variables have been masked or are simply not included in the Final Release public dataset. Some of these variables may be made available to analysts as restricted data. See our Web site for details.

Names, addresses, days of birth, information on geographical relocation and similar variables are not included in publicly released files. See also [Variables Not Included Publicly Released](#).

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. Denominations have been grouped. 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 "F972M" is the masked version of variable " F972". In all, the following variables have been masked for confidentiality.

F972M	A2A.REGION - US BORN	H98A_R
F976M	A2B.REGION - NON US BORN	H98A_R
F990M	A3C.R HIGH DEGREE - MASKED	H98A_R
F1003M1M	A7A.TYPE HISPANIC - MASKED	H98A_R
F1005M	A8.R RACE - MASKED	H98A_R
F1035M	A27.WHERE LIVE WHEN IN SCH - REGION	H98A_R
F1053M	A36A.DENOMINATION - MASKED	H98A_R
F997HM	A4F.FATHER USUAL OCCUPATION - HRS MASKED	H98A_R
F997AM	A4F.FATHER USUAL OCCUPATION - AHD MASKED	H98A_R
F1243M1M	B19A-1. MENTION HEALTH-1 - MASKED	H98B_R
F1243M2M	B19A-1. MENTION HEALTH-2 - MASKED	H98B_R
F1243M3M	B19A-1. MENTION HEALTH-3 - MASKED	H98B_R
F731M	CS33.CURRENT RESIDENCE - REGION	H98CS_H
F736M	CS35.OTH RESIDENCE - REGION	H98CS_H
F720M	CS25B.FACILITY WHERE LIVE - REGION	H98CS_R
F1940M	D106.REGION THEY LIVE	H98D_H
F1953M	D115.REGION MOTHER LIVE	H98D_H
F1966M	D125.REGION FATHER LIVE	H98D_H
F2102M	D106-2.REGION THEY-IN-LAW LIVE	H98D_H
F2115M	D115-2.REGION MOTHER-IN-LAW LIVE	H98D_H
F2128M	D125-2.REGION FATHER-IN-LAW LIVE	H98D_H
F2010M	D143c.SIB/S-IN-LAW HIGHEST DEGREE-MASKED	H98D_SB
F2916M	F40B.REGION OF 2ND HOME	H98F_H
F3185HM	G23.PREVIOUS WAVE WORK DONE - HRS MASKED	H98G_R
F3187HM	G25.CURRENT OCC - HRS MASKED	H98G_R
F3253M	G41A.IND - MASKED	H98G_R
F3255HM	G42.OCC - HRS MASKED	H98G_R
F3650M	GG4.IND - MASKED	H98G_R
F3652HM	GG6.OCC - HRS MASKED	H98G_R
F3839M	GH5.IND - MASKED	H98G_R
F3841HM	GH6.OCC - HRS MASKED	H98G_R
F4054HM	GD12D.OCCUPATION - HRS MASKED	H98G_R
F4055M	GD12E.INDUSTRY - MASKED	H98G_R
F4318HM	GJ40.OCCUPATION - HRS MASKED	H98G_R
F4319M	GJ41.INDUSTRY - MASKED	H98G_R
F4359HM	GJ61.OCCUPATION - HRS MASKED	H98G_R
F4360M	GJ62.INDUSTRY - MASKED	H98G_R
F4394HM	GJ80.OCCUPATION - HRS MASKED	H98G_R
F4395M	GJ81.INDUSTRY - MASKED	H98G_R
F4408HM	GJ90.PREVIOUS WAVE WORK DONE -HRS MASKED	H98G_R
F4409M	GJ91.INDUSTRY - MASKED	H98G_R
F4018M1M	GD1A.CAUSE PROB - MASKED	H98G_R
F4018M2M	GD1A.CAUSE PROB - MASKED	H98G_R
F4023M1M	GD1F.CAUSE PROB - MASKED	H98G_R
F4023M2M	GD1F.CAUSE PROB - MASKED	H98G_R
F4202M1M	GJ1A.CAUSE PROB - MASKED	H98G_R

F4202M2M	GJ1A.CAUSE PROB - MASKED	H98G_R
F4207M1M	GJ1F.CAUSE PROB - MASKED	H98G_R
F4207M2M	GJ1F.CAUSE PROB - MASKED	H98G_R
F4278M	GJ29A.CAUSE PROB - MASKED	H98G_R
F3185AM	G23.PREVIOUS WAVE WORK DONE - AHD MASKED	H98G_R
F3187AM	G25.CURRENT OCC - AHD MASKED	H98G_R
F3255AM	G42.OCC - AHD MASKED	H98G_R
F3652AM	GG6.OCC - AHD MASKED	H98G_R
F3841AM	GH6.OCC - AHD MASKED	H98G_R
F4054AM	GD12D.OCCUPATION - AHD MASKED	H98G_R
F4318AM	GJ40.OCCUPATION - AHD MASKED	H98G_R
F4359AM	GJ61.OCCUPATION - AHD MASKED	H98G_R
F4394AM	GJ80.OCCUPATION - AHD MASKED	H98G_R
F4408AM	GJ90.PREVIOUS WAVE WORK DONE -AHD MASKED	H98G_R
F4599M	H10B.MOVE TO ANOTHER REGION	H98H_R
F5854M	N63c.REGION DIED	H98N_R

7A3c. Variables with Alternative Format

A few variables have been derived. They have an "A" extension and are listed below. They appear in the codebook following the associated variable.

Much of Section A was asked only of new respondents. For your convenience, past wave data were merged for a few variables for previously interviewed respondents. For example, variable "F970A" contains data from preload variables F29_1-F29_3 and F82_1-F82_3 and coversheet variable F490 combined with data from F970.

F968A	A1.BIRTHDATE MONTH - COMBINED	H98A_R
F970A	A1B.BIRTHDATE YEAR - COMBINED	H98A_R
F987A	A3.R EDUCATION - COMBINED	H98A_R
F1002A	A7.HISPANIC - COMBINED	H98A_R
F1005A	A8.R RACE - COMBINED MASKED	H98A_R
F1156A	B7. HEART CONDITION	H98B_R
F454A	HRS98 VERSION	H98CS_R
F2639A	HELPER RELATIONSHIP COMBINED SOURCE	H98E_HP
F11A	HHMEM REL TO IDFM - UPDATED - CORRECTED	H98PR_MC
F6697A	VQ FINAL VERSION	H98T_R

7A3d. 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 F3667 -- GG14.EARNINGS WHEN STARTED, and F3668 --GG14.PER where F3667 is a dollar amount and F3668 is a time period. In the codebook you'll find the question text with the first of these two variables, F3667, and a line "AMOUNT: PER:". The second of these two variables, F3668, includes no question text in the codebook. See the previous variable, F3667, for the question text.

Similarly, through out the codebook you'll find questions that asked about the date of an event that generated two variables, month and year. For example, the question A40a asking "In what month and year were you married?" followed in the codebook with "MONTH: YEAR:" generated two variables, F1061 -- A40A.POST W1 MARRIAGE MO and F1062 -- A40C.YEAR. In the codebook the first variable of this pair includes the question text; the second does not.

Also sometimes one CATI screen contained many questions. For instance, a series of questions, labeled B19, was asked about preventative health behaviors. The question text for all of the questions in the series appears with the first question. Subsequent questions contain no question text. The variable labels for the subsequent questions will provide information to know which question was asked.

7A4. Master Code

A master code file contains detailed codeframes used in several 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 1998 data were collected using the Surveycraft CAI (Computer Assisted Interviewing) program. Surveycraft uses program specification files to determine what the program is supposed to do. The file "H98SPC.TXT" is a basic translation of the HRS 1998 Surveycraft program specifications.

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 1998 HRS data were collected using a CAI program, a traditional hard-copy questionnaire was not produced as part of the data collection phase. However, for 1998 we have provided a version of the traditional box-and-arrow questionnaire to help document the asking sequence of the questions.

8. Additional Notes

Found here are miscellaneous additional notes regarding HRS 1998 Final Data Release, Version 2. If we become aware of additional issues, they will be posted on our Web site 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. We have noted this by including "UPDATED" in the label. However, all this can become confusing.

In addition to the "A" variables listed above, [Variables with Alternative Format](#), we recommend you use these variables for your analysis.

F548	NEW SP COUPLENESS - UPDATED	H98CS_R
F686	R SEX	H98CS_R
F693	SP GENDER	H98CS_R
F1071	A43A.MARITAL STATUS, ASSIGNED	H98A_R

8B. Selecting the Correct Variable for the Respondent from Preload Information

The preload section contains multiple instances of a number of variables. The information below will allow you to select the variable with the information the respondent.

For the variables in the list below, one of six variables contains information about the respondent. For instance, there are six variables that may contain information about the respondent's birth year, F29_1 to F29_3 for respondents in household 1 and F82_1 to F82_3 for respondents in household 2. In order to know which variable to use, you should use two additional variables from the coversheet section the first, F459, to determine whether to use the variables for household 1 or household 2, and the second, F453, to determine which of the three variables to use.

F459 - WHICH HH

1. Household 1
2. Household 2

F453 - R INDEX

1. OLD R (1ST SLOT)
2. OLD R (2ND SLOT)
3. NEW R (3RD SLOT)

Q19_1-Q19_3	HH1 R INDEX
Q20_1-Q20_3	HH1 PERSON NUMBER - UPDATED
Q21_1-Q21_3	HH1-R PERSON NUMBER SEED
Q22_1-Q22_3	HH1 RANDOM WORD/VOCAB LIST - UPDATED
Q23_1-Q23_3	HH1 - R`S CURRENT S/P RANDOM WORD/VOCAB
Q24_1-Q24_3	HH1 - R`S NEW S/P RANDOM WORD/VOCAB LIST
Q25_1-Q25_3	HH1 WHICH COHORT
Q26_1-Q26_3	HH1 R PREV WAVE IW
Q29_1-Q29_3	HH1 R YR BIRTH - UPDATED
Q30_1-Q30_3	HH1 R MO BIRTH - UPDATED
Q31_1-Q31_3	HH1 R DY BIRTH - UPDATED
Q32_1-Q32_3	HH1: R ORIG SAMP MEMB
Q33_1-Q33_3	HH1: R AGE ELIG
Q34_1-Q34_3	HH1-R COGNITIVE PROXY SEED - UPDATED
Q35_1-Q35_3	HH1 R SEX - UPDATED
Q36_1-Q36_3	HH1 R MARITAL STAT
Q37_1-Q37_3	HH1 R COUPLENESS - UPDATED
Q38_1-Q38_3	HH1 S/P DIE DURING STUDY
Q39_1-Q39_3	HH1 R`S HH INTERVIEWED IN PRIOR WAVE(S)
Q40_1-Q40_3	HH1 R SAME SPOUSE/PARTNER - UPDATED
Q41_1-Q41_3	HH1 PREV WAVE FAMILY/NONFAM - UPDATED
Q42_1-Q42_3	HH1 PREV WAVE FINAN/NONFIN - UPDATED
Q43_1-Q43_3	HH1 R WHICH HOUSEHOLD - UPDATED
Q44_1-Q44_3	HH1 1ST/2ND IW IN HH - UPDATED

Q45_1-Q45_3	HH1 R RELATION TO R - UPDATED
Q46_1-Q46_3	HH1 R LIVING - UPDATED
Q47_1-Q47_3	HH1 R IN NURSING HOME - UPDATED
Q72_1-Q72_3	HH2 R INDEX
Q73_1-Q73_3	HH2 R PERSON NUMBER - UPDATED
Q74_1-Q74_3	HH2 R PERSON NUMBER SEED
Q75_1-Q75_3	HH2 RANDOM WORD/VOCAB LIST - UPDATED
Q76_1-Q76_3	HH2 - R`S CURRENT S/P RANDOM WORD/VOCAB
Q77_1-Q77_3	HH2 - R`S NEW S/P RANDOM WORD/VOCAB LIST
Q78_1-Q78_3	HH2 WHICH COHORT
Q79_1-Q79_3	HH2 R PREV WAVE IW - UPDATED
Q82_1-Q82_3	HH2 BIRTH YR - UPDATED
Q83_1-Q83_3	HH2 R MO BIRTH - UPDATED
Q84_1-Q84_3	HH2 R DY BIRTH - UPDATED
Q85_1-Q85_3	HH2: R ORIG SAMP MEMB
Q86_1-Q86_3	HH2: R AGE ELIG
Q87_1-Q87_3	HH2-R COGNITIVE PROXY SEED - UPDATED
Q88_1-Q88_3	HH2 R SEX - UPDATED
Q89_1-Q89_3	HH2 R MARITAL - UPDATED
Q90_1-Q90_3	HH2 R COUPLENESS - UPDATED
Q91_1-Q91_3	HH2 S/P DIED DURING STUDY
Q92_1-Q92_3	HH2 R`S HH INTERVIEWED IN PRIOR WAVE(S)
Q93_1-Q93_3	HH2 R SAME SPOUSE/PARTNER - UPDATED
Q94_1-Q94_3	HH2 PREV WAVE FAMILY/NONFAM - UPDATED
Q95_1-Q95_3	HH2 PREV WAVE FINAN/NONFIN - UPDATED
Q96_1-Q96_3	HH2 R WHICH HH - UPDATED
Q97_1-Q97_3	HH2 R 1ST/2ND IW IN HH - UPDATED
Q98_1-Q98_3	HH2 R RELATION TO R - UPDATED
Q99_1-Q99_3	HH2 R LIVING - UPDATED
Q100_1-Q100_3	HH2 R IN NURSING HOME - UPDATED

For the variables listed below, one of three variables contains information about the respondent. For instance, there are three variables, F178_1 to F178_3, that may contain information about the respondent's previous wave coupleness. You should use F453 to determine which of the three variables to use.

Q160_1-Q160_3	VIRGIN R INDEX
Q161_1-Q161_3	VIRGIN R PERS NUM
Q162_1-Q162_3	VIRGIN R PERSON NUMBER SEED
Q163_1-Q163_3	VIRGIN RANDOM WORD/VOCAB LIST
Q164_1-Q164_3	VIRGIN R`S CURRENT S/P RANDOM WORD/VOCAB
Q165_1-Q165_3	VIRGIN R`S NEW S/P RANDOM WORD/VOCAB LIS
Q166_1-Q166_3	VIRGIN WHICH COHORT
Q167_1-Q167_3	VIRGIN PREV WAVE EVER IW
Q170_1-Q170_3	VIRGIN R PREV WAVE YR BIRTH
Q171_1-Q171_3	VIRGIN R MO BIRTH
Q172_1-Q172_3	VIRGIN R DY BIRTH
Q173_1-Q173_3	VIRGIN R ORIG SAMP MEMB
Q174_1-Q174_3	VIRGIN R AGE ELIG
Q175_1-Q175_3	VIRGIN- R COGNITIVE PROXY SEED
Q176_1-Q176_3	VIRGIN R PREV WAVE SEX
Q177_1-Q177_3	VIRGIN R MARITAL STAT
Q178_1-Q178_3	VIRGIN R COUPLENESS
Q179_1-Q179_3	VIRG S/P DIED AND NOT REPLACED DURING ST
Q180_1-Q180_3	VIRGIN R`S HH INTERVIEWED IN PRIOR WAVE(

Q181_1-Q181_3	VIRGIN R SAME SPOUSE/PARTNER
Q182_1-Q182_3	VIRGIN PREV WAVE FAMILY/NONFAM
Q183_1-Q183_3	VIRGIN PREV WAVE FINAN/NONFIN
Q184_1-Q184_3	VIRGIN R WHICH HOUSEHOLD
Q185_1-Q185_3	VIRGIN R 1ST/2ND IW IN HH
Q186_1-Q186_3	VIRGIN R REL TO R
Q187_1-Q187_3	VIRGIN R LIVING
Q188_1-Q188_3	VIRGIN R IN NURSING HOME

For the variables listed below, one of two variables contains information about the respondent. For instance, there are two variables, F50 and F103, that may contain information about the respondent's previous wave second residence. You should use F459 to determine which of the two variables to use.

Q50	HH1 PREV WAVE HAS 2ND RESIDENCE
Q53	HH1 2ND OWN RESIDENCE - UPDATED
Q56	HH1 R MOVED - UPDATED
Q103	HH2 PREV WAVE HAS 2ND RESIDENCE
Q106	HH2 2ND OWN RESIDENCE - UPDATED
Q109	HH2 1ST R MOVED - UPDATED

8C. Households with No or Two Coversheet or Family or Financial Respondents

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

8C1. Households with No Coversheet or Family or Financial Respondent

Household records are provided for all households. Twenty-four households had no coversheet respondent; 174 households had no family respondent; 135 households had no financial respondent. The household records for these households contain null values for the missing information.

Households missing a coversheet or family or financial respondent can be identified, respectively, by values of

FPN_CS - 1998 COVERSHEET RESP PERSON NUMBER
 FPN_FAM - 1998 FAMILY RESP PERSON NUMBER
 FPN_FIN - 1998 FINANCIAL RESP PERSON NUMBER

in the household record.

Some households had two respondents, neither of whom, in error, was designated as a family or financial respondent. There were 18 households with two respondents neither of whom was a family respondent; there were three households with two respondents neither of whom was a financial respondent.

Of these respondents the one whose PN appears as the FPN_FAM or FPN_FIN and the one whose PN appears as the FPN_NFAM or FPN_NFIN in the household record is indicated by the values of

FFINR - 1998 WHETHER FINANCIAL RESPONDENT

FFAMR - 1998 WHETHER FAMILY RESPONDENT

in the respondent record.

8C2. Households with Two Family or Financial Respondents

Four hundred and seventy-seven households had two coversheet respondents; eleven households had two family respondents; ten households had two financial respondents. The responses of one of these two respondents were selected for inclusion in the household-level files.

The respondents whose responses were used in the household records as well as those whose responses were not used are indicated, respectively, by the values of

FCSR - 1998 WHETHER COVERSHEET RESPONDENT
FFINR - 1998 WHETHER FINANCIAL RESPONDENT
FFAMR - 1998 WHETHER FAMILY RESPONDENT

in the respondent record.

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. These series of questions are included in the version 2, HRS 1998 Final Data Release for all of the unfolding bracket sequences.

In some versions of previous waves, rather than providing all of the unfolding bracket questions, we provided just 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.

More recently, and for this data release, instead of creating these summary bracket variables, imputations for most unfolding bracket variables have been or will be derived and provided separately. See the section [Imputation File](#) for more details about the contents of this file.

8E. 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. The coded answer was changed if it was determined that the comment changed the substance of the recorded answer. Occasionally additional codes were added to an existing codeframe. The text of added codeframes appears in the codebook in upper and lower case. Double reporting of income or assets in Section J was corrected wherever possible.

When a change was made as a result of the comment review process, inconsistencies with subsequent variables may result. A limited number of changes for consistency were made to immediately subsequent variables within the section. The INAP text in the codebook does not include codeframes added in the comment review process. No consistency changes were made to variables in other sections. If any comment change affected a subsequent branch-point, the branch-point was not changed.

8F. Modules for the 1998 Data Collection

There were limited modules in HRS 1998, due to the addition of two new cohorts and the merger of the original HRS and AHEAD cohorts. The 1998 modules were asked primarily of the AHEAD sample members who were asked corresponding modules in previous waves.

Module 1: AHEAD 1993 ADL Questions. Based on AHEAD 1995, wave 2, module 1. Most respondents who were asked to do this module in 1998 were asked to do the same module in AHEAD 1995. This module contains questions about difficulty and the use of equipment and help in activities of daily living. These questions were part of the core questions in AHEAD 1993.

Module 3: ADL Measures Used for the Longitudinal Studies of Aging (LSOA). Based on AHEAD 1993, wave 1, module 3. Most respondents who were asked to do this module in 1998 were asked to do the same module in both AHEAD 1993 and AHEAD 1995. This module contains the ADL questions used in AHEAD 1993 and AHEAD 1995, which were in turn based on those proposed for (and subsequently used in) the second LSOA.

Module 4: ADL Measures Used for the National Long-Term Care Study (NLTC). Based on AHEAD 1993, wave 1, module 4. Most respondents who were asked to do this module in 1998 were asked to do the same module in AHEAD 1993; this was not a module in AHEAD 1995.

Module 5: The 1990 Census ADL questions. Based AHEAD 1993, wave 1, module 5. Slightly less than half of the respondents who were asked to do this module in 1998 were asked to do the same module in AHEAD 1993 (most of the rest were asked to do module 6 in AHEAD 1993); this was not a module in AHEAD 1995.

Also included in the module section, although not actually a module, is a section labeled module 10. It is a section of the interview that is administered to sample persons for whom proxy interviews were done, but are willing and able to do this section themselves.

Module 10: Cognitive Section for Proxy Interviewed Sample Members. The cognitive section which the interviewer asked the sample member to do when a proxy interview had been done on their behalf because of cognitive problems.

8G. Two Versions of Masked Occupation Code

For AHEAD data collections in 1993 and 1995, occupation was coded using the SRC 2-Digit Occupation Code (adapted from the 1980 U. S. Census Occupation Code) and masked for public release. For HRS data collections in 1992, 1994 and 1996, occupation was coded using the 3-digit 1980 U. S. Census Occupation Code and masked for public release. The AHEAD and HRS masked occupation codeframes are not the same. The numbers in parenthesis in the two codeframes below are the 1980 U. S. Census Occupation Codes that map into the code category.

For 1998, the first year that the AHEAD and HRS sub-samples were merged for data collection, we are releasing occupation masked in two ways for the AHEAD sub-sample - using both the AHEAD and HRS masking schemes. Variables masked in the AHEAD tradition are named with a "AM" at the end of the variable name and variables masked in the HRS tradition are named with a "HM" at the end of the variable name, e.g., "F3185AM" and "F3185HM". In subsequent years we will use just the HRS masking scheme for all sub-samples. See our Web site for more detailed information about how occupation and industry were coded in each wave of the study.

8H. Variables Not Included Publicly Released

Some variables have not been included in the public release data for reasons of confidentiality. They may be made available to analysts as restricted data. See our Web site for details.

31	HH1 R DY BIRTH
84	HH2 R DY BIRTH
172	VIRGIN R DY BIRTH
489	CS2a2.DAY BORN
493	R DAY BORN
496	R DAY BORN
1029	A26a.HAVE YOU ALWAYS LIVED AT RES
1031	A26c.YEARS LIVED MAIN HU AREA
1032	A26c.SINCE YEAR LIVE MAIN HU
1138	B5g. CANCER START
4188	GD131b1.DAY
4552	GJ131b.DATE OF INJURY - DAY
5855	N63d.CAUSE OF DEATH

8I. Detailed Comparison of Variables in Version 1 and Version 2

Some variables previously included in the respondent-level coversheet file are now included in a household-level coversheet file instead. Version 2 also contains some new variables. A few open-ended questions were not coded during the initial round of coding; these variables were not included in version 1 of the HRS 1998 Final Core Data Release but are included in version 2.

720	CS25b.FACILITY WHERE LIVE
731	CS33.CURRENT RESIDENCE
736	CS35.OTH RESIDENCE CITY/STATE
1243	B19. OTHER HEALTH
4023	GD1f. CAUSE PROB
4054	GD12d.OCCUPATION
4055	GD12e.INDUSTRY
4318	GJ40.OCCUPATION
4319	GJ41.INDUSTRY
4359	GJ61.OCCUPATION
4360	GJ62.INDUSTRY
4394	GJ80.OCCUPATION
4395	GJ81.INDUSTRY
4408	GJ90.PREVIOUS WAVE WORK DONE
4409	GJ91.INDUSTRY

Also included in version 2 are a few variables that were not included in version 1.

1029	A26A.HAVE YOU ALWAYS LIVED AT RES
1031	A26C.YEARS LIVED MAIN HU AREA
1032	A26C.SINCE YEAR LIVE MAIN HU

A couple of variables were added or corrected which incorporate changes noted in data alerts for version 1. See the codebook for the corrected and uncorrected versions of these variables.

F11A	HHMEM REL TO IDFM - UPDATED - CORRECTED,
F1156A	B7. HEART CONDITION
F987A	A3.R EDUCATION - COMBINED

Also added in version 2 are a number of variables that, while not of particular substantive interest, are used in assignment statements or are referenced in INAP text. For a detailed comparison of variables in version 1 and version 2, see the [Data Alert section](#) at our Web site.

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.

For more details about the sample, see our Web site.

9B. Linking Respondents across Time

Respondent records from all prior waves may be linked to respondent records from 1998 by HHID and PN. The sub-household identifiers, ASUBHH for 1992, BSUBHH for 1993, W2SUBHH for 1994, DSUBHH for 1995, ESUBHH for 1996 and FSUBHH for 1998 can be used to link cross-sectional household data with the cross-sectional respondent data.

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, to link 1998 other person records with 1996 other person records use HHID, ESUBHH, and OPN.

For some 1998 other person records this will result in two 1998 records matching one 1996 record. For example, in the example above where the couple with children divorces, the 1998 record where HHID is 89012, FSUBHH is 1, and OPN is 101 is the report about the child from the mother and the record where HHID is 89012, FSUBHH is 2, and OPN is 101 is the report about the same child from the father. The value of ESUBHH for each of these 1998 records would 0. Each of these 1998 records would match the 1996 record about the child before the couple divorced as the 1996 record would have a HHID of 89012, a ESUBHH of 0 and a OPN of 101.

For other 1998 other person records this will result in no 1998 record matching a 1996 record. For example, in the example above where the couple divorces and both split-off households have new members, the record where HHID is 90123, FSUBHH is 1, and OPN is 151 is for the ex-wife's cohabiting mother, while the record where HHID is 90123, FSUBHH is 2, and OPN is 151 is for the ex-husband's new stepchild. The value of ESUBHH for each of these 1998 records would 0. However neither of these 1998 records would match a 1996 record before the couple divorced because no 1996 record would have a HHID of 89012, a ESUBHH of 0 and a OPN of 151.

9D. Summary of Changes from Prior Waves Not New for 1998

Over the course of the study there have been a number of changes in key variables. There are no new changes of this type for 1998. However, since the analyst may be working with some or all of the earlier waves of data, changes from previous years are documented below.

9D1. 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.

- o All digits of an identifier matter. The character string "000031010" is not the same thing as the numeric string "31010".
- o When we begin to deal with complex respondent identifiers such as those in HRS and attempted to create one identification variable by combing several, we run the risk of inadvertently corrupting the identifier due to loss of precision as a side effect of treating it as a numeric value.

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. The Ahead Wave 1 (1993) data has recently been re-released with character identification variables.

Also released as character variables 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:

```
F1766      D01b.NON-RES CHILD LIVES NEAREST
F1771_1    D03bs.WHICH CHILD MARRIAGE CHANGE
F2024001   D146b.WHICH SIB HELPED
F2582      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 older HRS datasets

convert the numeric identifiers to character type. As of this writing, the 1992 and 1994 releases still include numeric identification variables.

- o Each software package has different ways of converting variable types in an existing dataset.

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

```
data out.w2a;
  *      set attrib values before SET statement to
        control order in output file;
  attrib
  CHHID label='HOUSEHOLD IDENTIFIER' length=$6 format=$char6.
  CHHID 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:\hrs1998\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:\hrs1998\spssfiles2\w2a.sav'
  / drop=hhid pn
  / rename=(chhid=hhid) (cpn=pn)
  / keep hhid pn all
  / map.
```

- o Alternatively, before reading in datasets, you can change the variable type in the data descriptor 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)

9D2. Identification Variable HHID Has One Extra Digit

A "0" was added to the front of the identification variable HHID for HRS entry cohort 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.

Similarly, a "20" was added to the front of the identification variable HHID for AHEAD entry cohort households, in order to make it possible to combine HRS and AHEAD sub-samples for the 1998 data collection. The change makes the old AHEAD four digit HHID six digits long. For example, old HHID "1234" becomes "201234".

You will want to be aware of these changes when preparing merges with previous years' data.

9D3. INAP Codes Stored as Blanks

Beginning with HRS 1995 data, INAP codes were stored in the ASCII data files as blanks. In 1992 and 1994, INAPs were assigned to a number, usually zero or a value ending in "96". In 1993, INAPs were assigned to a system missing value. If you are merging the 1998 data with data from 1992, 1993 or 1994, you will want to be aware of these differences.

10. Other Data Files

10A. Tracker File

We attempt to aid you in tracking the HRS sample longitudinally through use of our "tracker file". The tracker file is available at our Web site. The current version of the tracker file, the Cross-Wave: Tracker File (v.3.1), is based on information from the 2000 HRS Core (Final) (v.1.0) and previous years' interviews.

This tracker file contains a record for every individual who was interviewed one or more times during the period 1992 through 2000 as well as a record for the non-responding spouse or partner of such a person.

In addition to the primary identification variables HHID and PN, the tracker file contains basic demographic information as well as some basic information obtained from the National Death Index (NDI). In addition it contains basic information about the interview year such as whether a person gave an interview in a particular wave so that it can be used to understand the interview history of respondents. The tracker file also contains cross-sectional weights for each interview year 1992 through 2000 for both respondent and household level analysis.

10B. Imputation File

Imputed variables based on the HRS 1998 Final Core Data Release, version 1, may be found in HRS 1998 Imputation Release (Version 3.0), December 18, 2002. The files are available at our Web site. Also included in these files are four aggregate variables -- main home equity, and second home equity, total assets, and household income.

These files contain imputations based on the final release data, and provide imputations all bracket variables from all sections satisfying our ad hoc selection rules listed below.

- 1) The Bracket Rule: In general, all the variables with brackets will be imputed, and included in the imputation file.
- 2) The Income/Asset Component Rule: A variable without brackets will be imputed if it is 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 will not be imputed if there are no enough valid observations.

Section D Family Structure transfer-from-child-level, household-level, and transfer-to-child-level, Section E Health Care Utilization helper-level and respondent-level, Section F Housing household-level file, Section J Assets and Income household-level file, Section N Widowhood and Asset Changes household-level file, Section R Health Insurance respondent-level files are provided.

11. Obtaining the Data

11A. Downloading the Data

HRS data are available for free to researchers and analysts. In order to obtain Health and Retirement Study public release data, you must first register at our Web site and get a username and password. Your username and password are required to download any data files.

11B. Registration and Conditions of Use

By registering, you agree to the Conditions of Use governing access to Health and Retirement public release data. 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. For more information concerning privacy issues, please read our Privacy Statement available at our Web site.

As part of the data registration process, you agree 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 at 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 outline and to provide an approach to using the 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 new, as of 9/1/2002, URL 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-mail: 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. Examples of Merging

Many analyses require variables that appear 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.

What should be the level of the analysis file - will it have one record per respondent, one record per household or what? What are the variables of interest? Predetermining what variables are needed for an analysis allows you to subset files to include only the necessary variables, weights, and identification variables. Smaller files are more manageable. What identification variables will be required to merge the various files that contain variables needed for your analysis? Will the merge be a one-to-one matching of records or a one-to-many matching?

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.

13A. 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 21,384 records. A respondent-level output file with 21,384 respondent records results.

13B. 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 FSUBHH. This is a one-to-one match. Each input file contains 14,395 records. A household-level output file with 14,395 household records results.

13C. 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 FSUBHH. This is a one-to-many match (one household-to-many respondents). Household-level input files contain 4,395 records; respondent-level input files contain 21,384 records. A respondent-level output file with 21,384 respondent records results.

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

13D. 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 FFAMR - 1998 WHETHER FAMILY RESPONDENT to create the two groups of respondents. The first group will contain 14,224 records (174 households did not have a family respondent); the second group will contain 7,163 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 FSUBHH) to avoid overlap in the combined file.

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

Although the respondent groups contain fewer than 14,395 records, this is a one-to-one match. A household-level file with 14,395 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.

13E. 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 14,224 observations one for each family respondent. The file "nfamr" will have 7,163 observations one for each non-family respondent. The combined file "resp" will have 21,384 observations one for each respondent.

13F. SAS Examples

13F1. SAS Example of Respondent to Respondent Merge

```
proc sort
  data=in.h98a_r
  out=h98a_r(keep=hhid pn f968 f970);
  by hhid pn;
run;
proc sort
  data=in.h98b_r
  out=h98b_r(keep=hhid pn f1100 f1104);
  by hhid pn;
run;
proc sort
  data=in.h98c_r
  out=h98c_r(keep=hhid pn f1480 f1484);
  by hhid pn;
```

```

run;
data resp;
  merge h98a_r h98b_r h98c_r;
  by hhid pn;
run;

```

13F2. SAS Example of Household to Household Merge

```

proc sort
  data=in.h98f_h
  out=h98f_h(keep=hhid fsubhh f2740 f2741);
  by hhid fsubhh;
run;
proc sort
  data=in.h98j_h
  out=h98j_h(keep=hhid fsubhh f4628 f4630);
  by hhid fsubhh;
run;
data hhold;
  merge h98f_h h98j_h;
  by hhid fsubhh;
run;

```

13F3. SAS Example of Household to Respondent Merge

```

proc sort
  data=in.h98a_r
  out=h98a_r(keep=hhid fsubhh pn f968 f970);
  by hhid fsubhh;
run;
proc sort
  data=in.h98j_h
  out=h98j_h(keep=hhid fsubhh f4628 f4630);
  by hhid fsubhh;
run;
data resp;
  merge h98a_r h98j_h;
  by hhid fsubhh;
run;

```

13F4. SAS Example of Respondent to Household Merge

```

* file of family respondents;
proc sort
  data=in.h98a_r(keep=hhid fsubhh ffamr pn f968 f970
  rename=(ffamr=ffamra pn=pna f968=f968a f970=f970a)
  where=(ffamra in(1 2 3 4)))
  out=h98a1;
  by hhid fsubhh;
run;
* file of non-family respondents;
proc sort
  data=in.h98a_r(keep=hhid fsubhh ffamr pn f968 f970
  rename=(ffamr=ffamrb pn=pnb f968=f968b f970=f970b)
  where=(ffamrb not in(1 2 3 4)))
  out=h98a2;
  by hhid fsubhh;
run;
* variables from a household-level file;

```

```

proc sort
  data=in.h98j_h(keep=hhid fsubhh f4628 f4630)
  out=h98j;
  by hhid fsubhh;
run;
* household-level file with information from two respondents;
data hhold;
  merge
  h98a1 h98a2 h98j;
  by hhid fsubhh;
run;

```

13F5. 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.h98d_h;
  * keep id variables;
  keep hhid pn;
  * keep parent variables;
  keep f1906 f1907;
  * assign person number for family r;
  pn=fpn_fam;
  attrib pn label='PERSON NUMBER' format=$char3.;
  * output records for households with family R;
  if pn ne '';
run;
* select parent variables from household file for non-family r;
data nfamr;
  set in.h98d_h;
  * keep id variables;
  keep hhid pn;
  * keep parent variables;
  keep f1906 f1907;
  * copy in-law variables in to output variable location;
  f1906=f2068;
  f1907=f2069;
  * assign person number for family r;
  pn=fpn_nfam;
  attrib pn label='PERSON NUMBER' format=$char3.;
  * output records for households with non-family R;
  if pn ne '';
run;
* concatenate files;
data resp;
  set famr nfamr;
run;

```

13G. Stata Examples

13G1. Stata Example of Respondent to Respondent Merge

```

set prefix "hral998"
* This is an optional statement
* Make sure all the data files are in directory "hral998"

use HHID PN F1100 F1104 using h98b_r
sort HHID PN

```

```

save tmp1, replace

use HHID PN F1480 F1484 using h98c_r
sort HHID PN
save tmp2, replace

use HHID PN F968 F970 using h98a_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

```

13G2. *Stata Example of Household to Household Merge*

```

use HHID FSUBHH F2740 F2741 using h98f_h
sort HHID FSUBHH
save tmp, replace

use HHID FSUBHH F4628 F4630 using h98j_h
sort HHID FSUBHH
merge HHID FSUBHH using tmp
drop _m
sort HHID FSUBHH
save hhold, replace

```

13G3. *Stata Example of Household to Respondent Merge*

```

use HHID FSUBHH F4628 F4630 using h98j_h
sort HHID FSUBHH
save tmp1, replace

use HHID PN FSUBHH F968 F970 using h98a_r
sort HHID FSUBHH
merge HHID FSUBHH using tmp1
tab _m
drop _m
save resp, replace

```

13G4. *Stata Example of Respondent to Household Merge*

```

use HHID FSUBHH F4628 F4630 using h98j_h
sort HHID FSUBHH
save tmp1, replace

use HHID FSUBHH FFAMR PN F968 F970 using h98a_r
keep if FFAMR==1|EFAMR==2|FFAMR==3|EFAMR==4
rename FFAMR FFAMRa
rename PN PNa
rename F968 F968a
rename F970 F970a
sort HHID FSUBHH
save tmp2, replace

use HHID FSUBHH FFAMR PN F968 F970 using h98a_r
keep if FFAMR~=1 & FFAMR~=2 & FFAMR~=3 & FFAMR~=4

```

```

rename FFAMR FFAMRb
rename PN PNb
rename F968 F968b
rename F970 F970b
sort HHID FSUBHH
merge HHID FSUBHH using tmp2
tab _m
drop _m
sort HHID FSUBHH
merge HHID FSUBHH using tmp1
tab _m
drop _m
sort HHID FSUBHH
save hhold, replace

```

13G5. Stata Example of Respondent File with Parent Information from a Household-level File

```

* select parent variables from household file for family r
use h98d_h, clear

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

* keep ID variables and parent variables of interest
keep HHID PN F1906 F1907

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

* select parent variables from household file for non-family r
use h98d_h, clear
drop F1906 F1907

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

* copy in-law variables in to output variable location
gen F1906=F2068
gen F1907=F2069

* keep ID variables and parent variables of interest
keep HHID PN F1906 F1907

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

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

```

13H. SPSS Examples

13H1. SPSS Example of Respondent to Respondent Merge

/*-----

```

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

GET FILE 'c:\hrs1998\spss\h98a_r.sav'
      /keep=hhid pn f968 f970.
SORT CASES BY hhid pn.
save outfile='c:\temp\h98a_r.sav'.

GET FILE 'c:\hrs1998\spss\h98b_r.sav'
      /keep=hhid pn f1100 f1104.
SORT CASES BY hhid pn.
save outfile='c:\temp\h98b_r.sav'.

GET FILE 'c:\hrs1998\spss\h98c_r.sav'
      /keep=hhid pn f1480 f1484.
SORT CASES BY hhid pn.
save outfile='c:\temp\h98c_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\h98a_r.sav'
      /FILE='c:\temp\h98b_r.sav'
      /FILE='c:\temp\h98c_r.sav'
      /BY hhid pn.
EXECUTE.

save outfile='c:\hrs1998\spss\resp.sav'.

```

13H2. SPSS Example of Household to Household Merge

```

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

GET FILE 'c:\hrs1998\spss\h98f_h.sav'
      /keep=hhid fsubhh f2740 f2741.
SORT CASES BY hhid fsubhh.
save outfile='c:\temp\h98f_h.sav'.

GET FILE 'c:\hrs1998\spss\h98j_h.sav'
      /keep=hhid fsubhh f4628 f4630.
SORT CASES BY hhid fsubhh.
save outfile='c:\temp\h98j_h.sav'.

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

MATCH FILES
      /FILE='c:\temp\h98f_h.sav'
      /FILE='c:\temp\h98j_h.sav'
      /BY hhid fsubhh.
EXECUTE.

```

```
save outfile='c:\hrs1998\spss\hhold.sav'.
```

13H3. SPSS Example of Household to Respondent Merge

```
/*-----  
/*    sort each input dataset by HHID FSUBHH  
/*    subset variables  
  
GET FILE 'c:\hrs1998\spss\h98a_r.sav'  
    /keep=hhid fsubhh pn f968 f970.  
SORT CASES BY hhid fsubhh.  
save outfile='c:\temp\h98a_r.sav'.  
  
GET FILE 'c:\hrs1998\spss\h98j_h.sav'  
    /keep=hhid fsubhh f4628 f4630.  
SORT CASES BY hhid fsubhh.  
save outfile='c:\temp\h98j_h.sav'.  
  
/*-----  
/*    merge two datasets by HHID FSUBHH to create a respondent file  
/*    this is a one-to-many merge  
  
MATCH FILES  
    /FILE='c:\temp\h98a_r.sav'  
    /TABLE='c:\temp\h98j_h.sav'  
    /BY hhid fsubhh.  
EXECUTE.  
  
save outfile='c:\hrs1998\spss\resp.sav'.
```

13H4. SPSS Example of Respondent to Household Merge

```
/*-----  
/*    file of family respondents  
  
GET FILE 'c:\hrs1998\spss\h98a_r.sav'  
    /keep=hhid fsubhh ffamr pn f968 f970.  
USE ALL.  
SELECT IF(ffamr = 1 or ffamr = 2 or ffamr = 3 or ffamr=4).  
sort cases by hhid fsubhh.  
rename variables ffamr=ffamra.  
rename variables pn=pna.  
rename variables f968=f968a.  
rename variables f970=f970a.  
EXECUTE .  
  
save outfile='c:\temp\h98a1.sav'.  
  
/*-----  
/*    file of non-family respondents  
  
GET FILE 'c:\hrs1998\spss\h98a_r.sav'  
    /keep=hhid fsubhh ffamr pn f968 f970.  
sort cases by hhid fsubhh.  
USE ALL.  
SELECT IF(ffamr <> 1 and ffamr <> 2 and ffamr <> 3 and ffamr<>4).  
rename variables ffamr=ffamrb.  
rename variables pn=pnb.
```

```

rename variables f968=f968b.
rename variables f970=f970b.
EXECUTE .
save outfile='c:\temp\h98a2.sav'.

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

GET FILE 'c:\hrs1998\spss\h98j_h.sav'
    /keep=hhid fsubhh f4628 f4630.
SORT CASES BY hhid fsubhh.
save outfile='c:\temp\h98j.sav'.

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

MATCH FILES
    /FILE='c:\temp\h98a1.sav'
    /FILE='c:\temp\h98a2.sav'
    /FILE="c:\temp\h98j.sav"
    /BY hhid fsubhh.
EXECUTE.

save outfile='c:\hrs1998\spss\hhold.sav'.

```

13H5. 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:\hrs1998\spss\h98d_h.sav'
    /keep=hhid fpn_fam f1906 f1907.

/*    assign person number for family r
STRING pn (A3).
COMPUTE pn = fpn_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= fpn_fam.

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

/* keep id variables and parent variables of interest
GET FILE 'c:\hrs1998\spss\h98d_h.sav'
    /keep=hhid fpn_nfam f1906 f1907.

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

```

```

/*      copy in-law variables in to output variable location
rename variables f1906= f2068.
rename variables f1907= f2069.
USE ALL.

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

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

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

save outfile='c:\hrs1998\spss\resp.sav'.

```

14. 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 reunited³.

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

14A. 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 12345 and a SUBHH of 0. One respondent has a PN of 010, the other a PN of 020.

³ 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.

At the time of the second cross-section the two respondents are still married, and each retains their HHID of 12345 and their SUBHH of 0 and his and her PN of 010 and 020, respectively.

14B. 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 23456 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 23456, 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.

14C. 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 34567 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 34567. The living respondent retains her SUBHH of 0; the deceased respondent is assigned a SUBHH of 3. If both respondents die, one is assigned a SUBHH of 3 and the other is assigned a SUBHH of 4. Each original respondent retains his and her PN of 010 and 020, respectively.

14D. Single Respondent Marries.

A respondent who has never been married is in the first cross-section. The respondent is assigned a HHID of 45678 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 45678 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.

14E. 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 46789 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 46789, 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 56789 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 56789. 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.

14F. 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 67890 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 67890, 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 67890 and the SUBHH of 7. Each original respondent retains his and her PN of 010 and 020, respectively.

14G. Married Couple with Children and Siblings.

At the time of the first cross-section, sample household with a HHID of 89012 contains two respondents assigned PNs of 010 and 020, respectively. Associated with the household are three children with OPNs of 101, 102, and 201, 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.

14H. Couple with Children and Siblings Divorces.

At the time of the first cross-section, sample household with a HHID of 89012 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 89012, but he is assigned a SUBHH of 1 and the she is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. The two children appear in both their father's SUBHH 2 and 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.

14I. 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 90123 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 90123, but she is assigned a SUBHH of 1 and he is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively. Her mother is assigned the HHID of 90123 and the SUBHH of 1 and an OPN of 151. His new spouse and new stepchildren are assigned the HHID of 90123 and the SUBHH of 2. His new spouse

is assigned an OPN of 011. His new stepchildren are assigned OPNs of 151 and 152.