

HEALTH AND RETIREMENT STUDY  
2002 Core  
Final, Version 2.0  
June 2006

Data Description and Usage



---

---

**TABLE OF CONTENTS**

---

---

**TABLE OF CONTENTS**..... **III**

**DATA DESCRIPTION AND USAGE** ..... **1**

**1. OVERVIEW**..... **1**

    1A. CHANGES INCLUDED IN VERSION 2.0..... 1

    1B. THE SAMPLE INTERVIEWED IN 2002..... 1

    1C. 2002 QUESTIONNAIRE SECTIONS..... 2

    1D. LEVELS OF FILES ..... 3

        1D1. *Household Level Files*..... 3

        1D2. *Respondent Level Files*..... 3

        1D3. *Sibling Level File*..... 3

        1D4. *Household Member and Child Level Files*..... 3

        1D5. *Helper Level File*..... 3

        1D6. *Transfer-to-Child-Level File*..... 3

        1D7. *Transfer-from-Child-Level-File*..... 4

        1D8. *Asset Level File*..... 4

**2. FILE NAMING CONVENTIONS**..... **4**

**3. DATA FILES**..... **5**

**4. IDENTIFICATION VARIABLES** ..... **6**

    4A. PRIMARY IDENTIFICATION VARIABLES ..... 6

        4A1. *HHID - Household Identifier*..... 6

        4A2. *HSubHH - 2002 Sub-household Identifier*..... 6

        4A3. *PN - Person Number*..... 6

        4A4. *OPN - Other Person Number*..... 6

        4A5. *HTC\_NDX or HFC\_NDX - 2002 Transfer Number*..... 6

        4A6. *Asset Level Files*..... 6

    4B. PRIMARY IDENTIFICATION VARIABLES FOR DATASETS AT EACH OF THE SEVEN LEVELS..... 7

    4C. SECONDARY IDENTIFICATION VARIABLES..... 7

    4D. DATASETS INCLUDING HSubHH AND OPN AS PRIMARY IDENTIFIERS ..... 8

**5. DISTRIBUTION FILES AND DIRECTORY STRUCTURE** ..... **8**

    5A. DISTRIBUTION FILES ..... 8

    5B. DIRECTORY STRUCTURE..... 9

**6. PROGRAM STATEMENTS**..... **9**

    6A. USING THE FILES WITH SAS..... 9

    6B. USING THE FILES WITH SPSS..... 9

    6C. USING THE FILES WITH STATA..... 10

**7. DOCUMENTATION** ..... **10**

    7A. CODEBOOK..... 10

        7A1. *Variable Names*..... 10

        7A2. *Master Code*..... 11

    7B. OTHER TYPES OF DOCUMENTATION..... 11

        7B1. *2000 HRS Final Release Data Description*..... 11

        7B2. *Box and Arrow Questionnaire*..... 12

        7B3. *2000 to 2002 Variable Cross-Reference*..... 12

**8. ADDITIONAL NOTES** ..... **12**

8A. HOUSEHOLDS WITH NO COVERSREEN OR FAMILY OR FINANCIAL RESPONDENTS, OR TWO .....	12
8A1. <i>Households with No Coverscreen or Family or Financial Respondent</i> .....	12
8A2. <i>Households with Two CoverScreen or Two Family or Financial Respondents</i> .....	12
8B. COGNITION MEASURES.....	13
8C. UNFOLDING BRACKET VARIABLES AND IMPUTATIONS.....	13
8CA. <i>'97. Data Not Available' in Result Variables</i> .....	14
8D. MODULES FOR THE 2002 DATA COLLECTION.....	15
8E. EXPLANATION OF THE DIFFERENCE BETWEEN TRACKER AND 2002 FINAL DATA .....	15
8F. CHANGES IN DATA BECAUSE OF RESPONDENT COMMENTS .....	16
8G. WIDTH CHANGES IN THE DATA .....	16
<b>9. OBTAINING THE DATA.....</b>	<b>17</b>
9A. REGISTRATION AND DOWNLOADING THE DATA.....	17
9B. CONDITIONS OF USE .....	17
9C. PUBLICATIONS BASED ON DATA .....	17
<b>10. IF YOU NEED TO KNOW MORE .....</b>	<b>17</b>
10A. HRS INTERNET SITE.....	17
10B. CONTACT INFORMATION .....	18
<b>APPENDIX .....</b>	<b>19</b>
<b>EXAMPLES OF SUB-HOUSEHOLD AND RESPONDENT PERSON NUMBER AND OTHER PERSON NUMBER ASSIGNMENTS .....</b>	<b>19</b>
A1. MARRIED COUPLE STAYS MARRIED.....	19
A2. COUPLE DIVORCES.....	19
A3. ONE OR BOTH RESPONDENTS DIE. ....	20
A4. SINGLE RESPONDENT MARRIES.....	20
A5. COUPLE DIVORCES, ONE RESPONDENT REMARRIES AND DIVORCES.....	21
A6. COUPLE DIVORCES AND MARRIES AGAIN. ....	21
A7. MARRIED COUPLE WITH CHILDREN AND SIBLINGS. ....	22
A8. COUPLE WITH CHILDREN AND SIBLINGS DIVORCES. ....	22
A9. COUPLE DIVORCES, ONE RESPONDENT REMARRIES, BOTH SPLIT-OFF HOUSEHOLDS HAVE NEW MEMBERS. ....	23

---

## Data Description and Usage

### 1. Overview

The 2002 HRS Core (Final, Version 2.0) data release consists of data obtained as part of the Health and Retirement Study (HRS), a national longitudinal study of the economic, health, marital, and family status, as well as public and private support systems, of older Americans. The National Institute on Aging provided funding (NIH U01 AG09740), with supplemental support from the Social Security Administration. The Institute for Social Research (ISR) Survey Research Center (SRC) at the University of Michigan conducted the survey.

By receiving the data, which have been freely provided, you agree to use them for research and statistical purposes only and make no effort to identify the respondents. In addition, you agree to send HRS a copy of any publications you produce based on the data. See [Obtaining the Data](#) for additional details.

#### 1A. Changes Included in Version 2.0

The 2002 HRS Core (Final, Version 2.0) contains an additional case (HHID 200146 PN 020), which was left out of the HRS 2002 Final Version 1.0 data. Although some of the data are missing for this case, keeping the case was deemed preferable to dropping it.

We have also incorporated several changes into the 2002 HRS Core Final (Version 2.0) that were noted in data alerts for Version 1.0. They include the following:

#### October 10, 2005

One case was deleted - HHID - 053846 PN - 011. It was determined that PN 010 and PN 011 were the same individual.

#### August 10, 2005

Respondent Person number (PN) was added to the file H02PR\_SB as a primary Identification Variable.

A user note was added to the codebook in Section P, variable HX078.

Vocabulary variables HRC161 HRC163 HRC165 HRC167 HRC169 were coded and corrected in section RC.

Corrections were made to the variable HX026M in the file H02PR\_H.

#### April 4, 2005

Missing data were recovered for variable HJ585 in file H02J-R.

October 25, 2004 - Missing data were recovered for six cases in file H02D\_R for variables HD182M1 and HD182M2. Missing data were recovered for 93 cases for HD183M1 through HD183M10. One duplicate line was deleted from the File H02E\_MC. One duplicate OPN in file H02PR\_MC was changed to a new value.

#### 1B. The Sample Interviewed in 2002

The data collection period for the 2002 interview was February 2002 through March 2003. As of 2002, the HRS sample was comprised of four sub-samples.

The first sub-sample, 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 their spouses or partners at the time of the initial interview in 1992 or at the time of any subsequent interview. The HRS sub-sample was interviewed in 1992 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 their spouses or partners at the time of the initial interview or at the time of any subsequent interview. The AHEAD sub-sample was interviewed in 1993-94, 1995-96, 1998 and every two years thereafter.

The War Baby (WB) sub-sample consists of people who were born in 1942 through 1947, were household residents of the conterminous U.S. in the spring 1992, who, at that time, did not have a spouse or partner born before 1924 or between 1931 and 1941, and were still household residents at the time of the first interview in 1998, and their spouses or partners at the time of the initial interview or at the time of any subsequent interview. The War Baby sub-sample was interviewed in 1998 and every two years thereafter.

The Children of the Depression (CODA) sub-sample consists of people who were born in 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, and their spouses or partners at the time of the initial interview or at the time of any subsequent interview. The Children of the Depression sub-sample was interviewed in 1998 and every two years thereafter.

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

#### 1C. 2002 Questionnaire Sections

The content of the 2002 data collection instrument is roughly equivalent to the content of the HRS 2000 instrument. However, in 2002 the data collection instrument required new programming due to a change in the software used to collect the data. Consequently, users who are familiar with the HRS will notice that the section headings and variable names have changed markedly between HRS 2000 and HRS 2002. Also, new sections "Asset Reconciliation" (Section U) and "Interviewer Observations" (Section IO) have been added to the 2002 data. The following table presents the sections of the questionnaire as they are in the HRS 2002 instrument versus their appearance in the HRS 2000.

2002 Section	Content	2000 Section
PR	Preload	PR
A	CoverScreen	CS
B	Demographics	A
C	Health	B
D	Cognition	PC/C
RC	Repeat Cognition	RC
E	Family Structure	D1
F	Parents/Sibs/Decisions	D2
G	Functional Limitations	E2
H	Housing	F
J	Employment	G
K	Last Job	GG
L	Job History	GH
M	Disability	GD/GJ
N	Health Services and Insurance	E1/R1
P	Expectations	H
Q	Assets and Income	J1
R	Asset Change	N1
S	Widowhood and Divorce	N2
T	Wills and Life Insurance	J2/R2
U	Asset Reconciliation	-
V	Modules	MD
W	Event History and Internet Use	EV
Y	Timestamps	TS
IO	Interviewer Observations	-

## 1D. Levels of Files

In the 2002 data collection instrument, most questions were asked of all respondents. Some questions were asked about the household. For two-respondent households, household level questions were asked of one respondent who was designated as the financial respondent, family respondent, or coverscreen respondent (the first respondent interviewed) on behalf of the entire household.

In addition to the familiar household-level and respondent-level files, the 2002 HRS Core (Final, Version 2.0), contains files at six other levels: household-member-and-child, sibling, helper, transfer-to-child, transfer-from-child, asset and pension.

### *1D1. Household Level Files*

Household-level files contain questions that were asked about the household of a designated coverscreen, financial, or family respondent. A coverscreen respondent answered family questions (section A) on behalf of the entire household; the coverscreen respondent may or may not be the family respondent. A family respondent answered family questions (section E) on behalf of the entire household, and a financial respondent answered household-level financial questions (sections H, Q, and R) 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 2002.

### *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 contain one record for each respondent or proxy who gave an interview in 2002.

### *1D3. Sibling Level File*

The sibling file consists of characteristics of the respondent's siblings. If a respondent had at least one living parent, he/she was asked a variety of questions about his/her siblings (Section F). The sibling file contains one record for each sibling of a respondent. In prior years, the designated family respondent reported on his/her own siblings and also on the siblings of his/her spouse or partner. The 2002 data are different from prior waves in that each respondent reports on their own parents and siblings. Sibling data are also stored in the preload section, H02pr\_sb.

### *1D4. Household Member and Child Level Files*

These files contain characteristics about household members and children. The information can come from Section A (coverscreen) or the family respondent in Section E and also in the preload section.

### *1D5. Helper Level File*

This file contains information provided by each respondent about helpers. A helper may be a person or organization that was reported by the respondent as providing help with ADLs or IADLs. The file contains one record for each helper. If a child helped both respondents in a two respondent household, the helper file will contain two records - one for mother's report of the child's helping her and one for father's report of the child's helping him.

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

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

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

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

### *1D8. Asset Level File*

The Asset file has one record per asset eligible for "Asset Reconciliation". The data in the asset level file were collected in section U (Asset Reconciliation). Section U compares the value of a select number of assets that were preloaded, and compares the preloaded value with the value given in section Q (Assets and Income), in order to determine if the value of an asset needs to be "reconciled" (i.e., are the two values too different from one another?). Section U data are the answers respondents gave for assets that appear to be discrepant on their preload and/or asset and income counterparts.

### *1D9. Pension Level File*

The pension level file consists of one record for each pension reported by respondents. The pension data in the pension level file (H02pr\_p) were collected in prior years, and these data would have been part of the Section GG (Last Job) and Section GH (Job History). They have been preloaded to gather some further information on past wave pension data.

## **2. File Naming Conventions**

Files are named beginning with "H02" for HRS 2002, followed by a letter (or two) designating the questionnaire section. A separator, "\_" and then one or two letters designating the level, follows the section letter designator. The following letters designate the level of the data files:

- 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
- FC for transfer-from-child-level
- A for asset level
- P for pension level

For example, the file H02A\_R includes variables from section A (coverscreen) at the respondent level. Alternatively, file H02U\_A includes variables from section U and the data are at the asset level.

The following extensions are used for the 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 the 36 data files for the 2002 HRS Core (Final, Version 2.0) data release. For example,

- H00A\_R.DA contains respondent data from section A,
- H00A\_R.SAS contains corresponding SAS program statements,
- H00A\_R.SPS contains corresponding SPSS program statements,
- H00A\_R.DO contains corresponding Stata DO statements,
- H00A\_R.DCT contains corresponding Stata dictionary statements, and
- H00A\_R.TXT contains the ASCII codebook.



### 3. Data Files

The 2002 HRS Core (Final, Version 2.0) data are distributed in 36 data files. The files are listed below along with the number of cases (N), number of variables (NV), and the primary identifiers (IDS). The records in the data files are sorted in order by these primary identifiers.

The 2002 HRS Core data are provided in ASCII format, with fixed-length records. Use associated SAS, SPSS or Stata program statements to read the data into the analysis package of your choice.

Household level files, IDS=HHID HSUBHH

H02PR_H	N=12,350	NV=32
H02A_H	N=12,350	NV=46
H02E_H	N=12,350	NV=109
H02H_H	N=12,350	NV=195
H02Q_H	N=12,350	NV=698
H02R_H	N=12,350	NV=112

Respondent level files, IDS=HHID PN

H02PR_R	N=18,167	NV=127
H02A_R	N=18,167	NV=28
H02B_R	N=18,167	NV=83
H02C_R	N=18,167	NV=193
H02D_R	N=18,167	NV=200
H02RC_R		
H02F_R	N=18,167	NV=136
H02G_R	N=18,167	NV=146
H02J_R	N=18,167	NV=1,395
H02K_R	N=18,167	NV=40
H02L_R	N=18,167	NV=117
H02M1_R	N=18,167	NV=272
H02M2_R	N=18,167	NV=453
H02N_R	N=18,167	NV=317
H02P_R	N=18,167	NV=71
H02S_R	N=18,167	NV=105
H02T_R	N=18,167	NV=85
H02V_R	N=18,167	NV=190
H02W_R	N=18,167	NV=23
H02Y_R	N=18,167	NV=52
H02RC_R		

Household member and child level files IDS=HHID HSUBHH OPN

H02PR_MC	N=70,128	NV=27
H02E_MC	N=67,971	NV=34

Sibling level files IDS=HHID HSUBHH OPN

H02PR_SB	N=14,129	NV=25
H02F_SB	N=6,196	NV=26

Transfer-to-child-level-file IDS=HHID HSUBHH OPN HTC\_NDX

H02E_TC	N=5,185	NV=20
---------	---------	-------

Transfer-from-child-level-file IDS=HHID HSUBHH OPN HFC\_NDX

H02E\_FC        N=925    NV=20  
Helper level file    IDS=HHID PN OPN  
  
H02G\_HP        N=4,686    NV=26  
Pension level file    IDS=HHID PN HKPNDX or HLPNDX  
  
H02PR\_P        N=565    NV=13  
Asset level file    IDS=HHID HSBUBHH TYPASST  
  
H02U\_A        N=1,478    NV=16

#### 4. Identification Variables

Identification variables for HRS 2002 are stored in character format.

##### 4A. Primary Identification Variables

Several variables, HHID, HSUBHH, PN, OPN, HTC\_NDX or HFC\_NDX and TYPASST are used in various combinations to uniquely identify the nine different level datasets that comprise this data release.

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

In the initial wave of data collection (in 1992 for the HRS sub-sample, in 1993 for the AHEAD sub-sample, and in 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. HSUBHH - 2002 Sub-household Identifier

In combination with HHID, HSUBHH uniquely identifies a household at the time of the 2002 data collection. Sub-household identifiers can be different at each wave. HSUBHH has one-digit. For more information, see [Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments](#).

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

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

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

In combination with HHID, HSUBHH, PN, and OPN uniquely identify another person - household member, child, sibling, or helper -- at the time of the 2002 data collection. OPN has three-digits. See [Datasets Including HSUBHH and OPN as Primary Identifiers](#) for additional information.

##### 4A5. HTC\_NDX or HFC\_NDX - 2002 Transfer Number

In combination with HHID and HSUBHH, HTC\_NDX or HFC\_NDX uniquely identifies transfer to or from a child or grandchild. HTC\_NDX (index for transfers to children) and HFC\_NDX (index for transfers from children) have one digit.

##### 4A6. Asset Level Files

Asset records (section U, data file H02U\_A) are uniquely identified by HHID, HSUBHH, and TYPASST

#### 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
- 2) PN PERSON NUMBER

Two identifiers uniquely identify records in the

- o household-level datasets:

- 1) HHID HOUSEHOLD IDENTIFIER
- 2) HSUBHH 2002 SUB-HOUSEHOLD IDENTIFIER

Three identifiers uniquely identify records in the

- o helper-level datasets:

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

Three identifiers uniquely identify records in the

- o household-member-and-child-level:

- 1) HHID HOUSEHOLD IDENTIFIER
- 2) HSUBHH 2002 SUB-HOUSEHOLD IDENTIFIER
- 3) OPN OTHER PERSON NUMBER

Four identifiers uniquely identify records in the

- o sibling-level datasets:

- 1) HHID HOUSEHOLD IDENTIFIER
- 2) OPN OTHER PERSON NUMBER
- 3) PN 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) HSUBHH 2002 SUB-HOUSEHOLD IDENTIFIER
- 3) OPN OTHER PERSON NUMBER
- 4) HTC\_NDX/HFC\_NDX 2002 TRANSFER NUMBER

Three identifiers uniquely identify records in the

- o Asset level dataset

- 1) HHID HOUSEHOLD IDENTIFIER
- 2) HSUBHH 2002 SUB-HOUSEHOLD IDENTIFIER
- 3) TYPASST TYPE OF ASSET

#### 4C. Secondary Identification Variables

In addition to the primary 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,
- o sibling level and
- o helper-level datasets:

HSUBHH	2002	SUB-HOUSEHOLD IDENTIFIER
GSUBHH	2000	SUB-HOUSEHOLD IDENTIFIER
HPN_SP	2002	SPOUSE/PARTNER PERSON NUMBER
HCSR	2002	WHETHER COVERSCREEN RESPONDENT
HFAMR	2002	WHETHER FAMILY RESPONDENT
HFINR	2002	WHETHER FINANCIAL RESPONDENT
HQNR	2002	QUESTIONNAIRE

#### Secondary Identification Variables for

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

GSUBHH	2000	SUB-HOUSEHOLD IDENTIFIER
HPN_CS	2002	COVERSCREEN RESP PERSON NUMBER
HPN_FAM	2002	FAMILY RESP PERSON NUMBER
HPN_FIN	2002	FINANCIAL RESP PERSON NUMBER
HPN_NCS	2002	NON-COVERSCREEN RESP PERSON NUMBER
HPN_NFAM	2002	NON-FAMILY RESP PERSON NUMBER
HPN_NFIN	2002	NON-FINANCIAL RESP PERSON NUMBER
HQNR	2002	QUESTIONNAIRE

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

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

Sometimes records with the same HHID and OPN will be separate reports about the same person. For example, see the example using 2000 data [Couple with Children and Siblings Divorces](#), where the couple with children divorces, the record where HHID is 89012, HSUBHH is 1, and OPN is 101 is the report about the child from the mother, and the record where HHID is 89012, HSUBHH 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, see the example using 2000 data [Couple Divorces, One Respondent Remarries, Both Split-off Households Have New Members](#). where the couple divorces and both split-off households have new members, the record where HHID is 90123, HSUBHH is 2, and OPN is 151 is for the ex-wife's mother, while the record where HHID is 90123, HSUBHH is 1, and OPN is 151 is for the ex-husband's new stepchild.

## 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, one .pdf file, and one .txt file, or the nine smaller files available individually for separate download. The combined file is h02core.zip.

The individual .zip files for separate download are:

Data file  
h02da.zip containing data files.

Program statement files  
h02sas.zip containing SAS data descriptors.  
h02sps.zip containing SPSS data descriptors.

h02sta.zip containing Stata data descriptors.

#### Documentation files

h02cb.zip containing the codebook.  
h02qn.zip containing the questionnaire.  
h02xref.txt - the HRS 2002 to HRS 2000 cross-reference file  
h02dd.pdf - this document.

#### 5B. Directory Structure

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

Directory	Contents
c:\hrs2002	Files downloaded from Web site
c:\hrs2002\codebook	Unzipped files from h02cb.zip
c:\hrs2002\data	Unzipped files from h02da.zip
c:\hrs2002\qnaire	Unzipped files from h02qn.zip
c:\hrs2002\sas	Unzipped files from h02sas.zip
c:\hrs2002\spss	Unzipped files from h02sps.zip
c:\hrs2002\stata	Unzipped files from h02sta.zip

Decompress the selected .zip files into the appropriate subdirectories. You will need about 191 MB of free space on your storage device to store the 36 .DA files.

## 6. Program Statements

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

The statement files are named beginning with the same prefix as the corresponding data file. For example, SAS statements in the file H02A\_R.SAS go with the H02A\_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:\h2002\sas" and the data file is located in "c:\h2002\data", you can run the file as is. A SAS system file (\*.SD2 or \*.SAS7BDAT) will be saved to directory "c:\h2002\sas".

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

#### 6B. Using the Files with SPSS

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

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

If the \*.SPS file is located in "c:\h2002\spss" and the data file is located in "c:\h2002\data", you can run the file as is. An SPSS system file (\*.SAV) will be saved to directory "c:\h2002\spss".

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

#### 6C. Using the Files with Stata

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

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

If the \*.DO and \*.DCT files are located in "c:\h2002\stata" and the data file is located in "c:\h2002\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.

Note that the variable names provided in the .DCT files are uppercase. If you prefer lower case variable names, you may wish to convert the .DCT files to lower case prior to use. You may do this by reading the .DCT file into a text or word processing program and changing the case. For instance in Microsoft Word, Edit, Select All, Format, Change Case, lowercase.

## 7. Documentation

There are several types of documentation available for use with the 2002 HRS Core (Final, Version 2.0) data release. These include a codebook, the 2002 box-and-arrow questionnaire, and a 2000 to 2002 variable cross-reference listing. In addition, 2002 variables have been added to the [Online Concordance](#).

### 7A. Codebook

The HRS 2002 Codebook is provided as a series of 36 ASCII text files, as well as a file containing all sections. There is a codebook file corresponding to each data file. Each variable has its own codebook entry. The format of the codebook is, for the most part, consistent with the codebooks of 1995, 1996, 1998, and 2000. The most notable difference is with the naming of variables in the 2002 data.

#### 7A1. Variable Names

Variable names begin with a letter designating the wave of data collection (H for 2002), followed by the section letter, and numbers after the section letter. For example, HC001 where H=2002, C=section C (physical health), 001 variable number. Variables from the preload section of the instrument will have either HX or HZ as prefix letters. The X indicates a variable that is updated by data collected in later sections of the questionnaire, whereas the Z indicates preloaded data that were not changed by subsequent answers to questions. For example:

```
HX007    RESP FAM/FIN TYPE - UPDATED
HZ007    PREV WAVE R FIN/FAM TYPE

HX065_R  COUPLENESS STATUS - UPDATED
HZ066_R  PREVIOUS WAVE COUPLENESS - INDIVIDUAL
```

#### 7A1a. Multiple-response and Looped Variables

There are two types of variables with multiple mention indicators. First are simple multiple mentions and second are multiple mentions within loops.

Simple multiple mention variables take the form: (wave prefix)+ (section letter) + (variable number) + (mention number). For example, HJ045M1 through HJ045M11 are 2002 variables from section J with 1 through 11 mentions.

Variable names for multiple mentions to questions within a loop take the form: (wave prefix) + (section letter) + (variable number) + (underscore) + (loop iteration) + (letter designating mention number). For example, HJ048\_1a, is a 2002 variable from section J, variable number 048 in the first iteration of the loop, and the first mention.

Simple loop variables (without a multiple mention) have an underscore ( \_ ) in their name and a suffix that designates the loop, e.g., HJ045\_1.

Null multiple mention variables and variables from null loops beyond the first mention or first loop are not included in the data. It is generally the case that one null multiple mention and one null loop was retained.

#### 7A1b. Masked Variables

To protect the confidentiality of the information that respondents 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.

Geographical locations are recoded to a level no more detailed than U.S. Census Region and Division. Data on the highest educational degree earned have been further grouped together to increase cell sizes. Industry and occupation codes have been recoded into a limited number of categories from the original three-digit U.S. Census code.

The names of variables that were masked for confidentiality end in the letter "M"; for example, variable HX026M (1ST ADDRESS STATE - MASKED) and variable HB024M (FATHER USUAL OCCUPATION - MASKED).

#### 7A1c. Other Specify Questions, Comments, and Open Ends

"Other Specify" and "Open End", or questions that are answered with text, are not included in Final releases. These variables will be coded and will be included in the Final release of the 2002 HRS data. Similarly, the Early release data do not include comments made by respondents in the course of the interview. For the Final release, HRS staff reviews these comments for selected questions, and the coded answer is changed if it is determined that the comment changed the substance of the recorded answer.

#### 7A2. 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, three additional types of documentation are available.

##### 7B1. 2000 HRS Final Release Data Description

The 2000 HRS Final Release Data Description document contains useful file merge examples and additional information about HRS data files in general.

### 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 2002 HRS data were collected using a CAI program, a traditional hard-copy questionnaire was not produced as part of the data collection phase. However, we have provided a version of the traditional box-and-arrow questionnaire to help document the asking sequence of the questions.

### 7B3. *2000 to 2002 Variable Cross-Reference*

As mentioned earlier, the names of the variables in the 2002 datasets are markedly different from the names of variables in prior waves. A variable cross-reference table has been assembled to facilitate finding variables of interest in the 2002 data. The cross-reference links the 2002 variable name, level, and label, with the 2000 equivalent(s). If a 2002 variable has a 2000 equivalent, a reference link appears in the 2002 codebook.

## 8. Additional Notes

Found here are miscellaneous additional notes regarding HRS 2002 Final Data Release, Version 1. If we become aware of additional issues, they will be posted on our Web site in the Data Alerts section.

### 8A. *Households with No Coverscreen or Family or Financial Respondents, or Two*

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 coverscreen 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; for a few other households we obtained interviews with two coverscreen or family or financial respondents.

#### 8A1. *Households with No Coverscreen or Family or Financial Respondent*

Household records are provided for all households. All households had at least one coverscreen respondent; 3 households had no family respondent; 31 households had no financial respondent. The household records for these households contain null values for the missing information.

Households missing a coverscreen or family or financial respondent can be identified, respectively, by values of "Blank. No coverscreen/family/financial respondent" in these variables:

GPN\_CS - 2002 COVERSCREEN RESP PERSON NUMBER  
GPN\_FAM - 2002 FAMILY RESP PERSON NUMBER  
GPN\_FIN - 2002 FINANCIAL RESP PERSON NUMBER

#### 8A2. *Households with Two CoverScreen or Two Family or Financial Respondents*

One hundred ninety-one households had two coverscreen respondents; 6 households had two family respondents; 9 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

GCSR - 2002 WHETHER COVERSCREEN RESPONDENT  
GFINR - 2002 WHETHER FINANCIAL RESPONDENT  
GFAMR - 2002 WHETHER FAMILY RESPONDENT



8B. Cognition Measures

There are two types of cognition questions that are asked in HRS. The first set of questions includes cognitive performance measures, and are only asked of self-respondents. The second set of questions are asked of proxy respondents. For a detailed description of these measures, please refer to the documentation report, <http://hrsonline.isr.umich.edu/docs/userg/dr-006.pdf>.

Cognition data for self-respondents are available in section D (Cognition). At the end of the interview when the cognition section is completed with a proxy reporter, the proxy is asked if the respondent him/herself is available to answer a few questions. If the proxy reporter agrees and the respondent is willing, then the respondent is asked the self-R cognition questions. Data obtained from these respondents are included in a separate section (RC) in order to separate them from responses obtained during self- rather than proxy-interviews.

8C. Unfolding Bracket Variables and Imputations

Typically, a series of unfolding bracket questions followed a lead-in question asking for an amount. If an actual amount was not given, a series of "unfolding" questions were asked. The manner in which the unfolding questions were programmed (Blaise) is different for 2002 data compared to the CAI (SurveyCraft) software used for 1993 through 2000. This change was transparent to the respondents, since exactly the same questions were asked with the new software as would have been asked with the old software; but it did have an implication for the data that were actually stored and also for the data that are released.

Instead of storing the response to each unfolding question, three summary variables were generated: the minimum and maximum values for the amount, given the answers to the unfolding questions, and if the last answer a respondent gave in an unfolding sequence was either "Don't Know" or "Refused," what that answer was. If the Respondent said "more than" to the unfolding question with the highest value, then the maximum value was stored as ten times that value.

For most analysts, those three variables (and in particular, the minimum and maximum of the possible range) will be sufficient for analyses. For any analyst who needs the more detailed information, it should be noted that the three variables, combined with the information about the unfolding questions provided in the box-and-arrow and codebook, are sufficient to allow the analyst to reconstruct the sequence of questions asked of any respondent, and the answers to each of those questions in many of the unfolding sequences.

For other sequences -- those in which respondents were randomly assigned to one of three "entry" points for the first unfolding question -- the analyst will also need to take into account a fourth variable (located in the preload sections) that specifies the entry point for each respondent. The following example shows the preload variable (HZ041) and the unfolding sequence that uses the random entry point from HZ041.

Example Random Entry Assignment Variable from Preload:

Preload Variable from the data file H02pr\_h:

=====

HZ041 UNFOLD ASSIGN - SELF EMPLOYMENT INCOME  
Section: PR Level: Household Type: Numeric Width: 1 Decimals: 0  
CAI Reference: B\_HouseHold.X041\_UnfSEmpInc\_V Ref 2000: G111

UNFOLD ASSIGN - SELF EMPLOYMENT INCOME  
.....  
4065 1. RANDOM ASSIGNMENT  
4124 2. RANDOM ASSIGNMENT  
4160 3. RANDOM ASSIGNMENT

1 Blank. INAP (Inapplicable)

=====  
Unfolding Series from section Q that uses HZ041 to assign respondents an  
entry point  
=====

HQ016 R INCOME FROM SELF EMPLOYMENT - MIN  
Section: Q Level: Household Type: Numeric Width: 6 Decimals: 0  
CAI Reference: BQ\_RIncome.Q016\_

Q016-Q018 Unfolding Sequence  
Question text: Does it amount to less than \$\_\_\_\_\_, more than  
\$\_\_\_\_\_, or what?

PROCEDURE: UNFM\_2up1down  
BREAKPOINTS: 5000, 10000, 25000, 100000

.....  
137 0. Value of Breakpoint  
7 5000. Value of Breakpoint  
8 5001. Value of Breakpoint  
6 10000. Value of Breakpoint  
26 10001. Value of Breakpoint  
2 25000. Value of Breakpoint  
25 25001. Value of Breakpoint  
1 100000. Value of Breakpoint  
17 100001. Value of Breakpoint  
12121 Blank. INAP (Inapplicable)

=====  
HQ017 R INCOME FROM SELF EMPLOYMENT - MAX  
Section: Q Level: Household Type: Numeric Width: 7 Decimals: 0  
CAI Reference: BQ\_RIncome.Q017\_

.....  
35 4999. Value of Breakpoint  
7 5000. Value of Breakpoint  
8 9999. Value of Breakpoint  
6 10000. Value of Breakpoint  
24 24999. Value of Breakpoint  
2 25000. Value of Breakpoint  
21 99999. Value of Breakpoint  
1 100000. Value of Breakpoint  
125 1000000. Value of Breakpoint  
12121 Blank. INAP (Inapplicable)

=====  
HQ018 R INCOME FROM SELF EMPLOYMENT - RESULT  
Section: Q Level: Household Type: Numeric Width: 2 Decimals: 0  
CAI Reference: BQ\_RIncome.Q018\_

.....  
27 98. DK (Don't Know)  
84 99. RF (Refused)  
12239 Blank. INAP (Inapplicable)

---

8CA. '97. Data Not Available' in Result Variables

The '97. Data Not Available' code was added to the Result variables in the unfolding sequences to account for inaccuracies in the skip patterns resulting

from the changes to the data from coding of respondent comments. The normal operations of the unfolding sequences is that the respondent is asked a number of questions to determine the maximum and minimum range of a value, when they respond with DK (Don't Know) or RF (Refuse) at the question where an amount is asked. However, sometimes a respondent gives an amount after the series had been skipped. The value variable is then coded as a DK (Don't Know), but the unfolding sequence remains blank. In cases such as this, they will have a '97. Data Not Available' value in the Result variable of the unfolding sequence.

#### 8D. Modules for the 2002 Data Collection

There are 11 modules for 2002 HRS (there is no module number 7). Topical areas of the modules include:

- Module 1: Self-Assessed Health Utilities
- Module 2: Willingness to Pay for Disease Prevention
- Module 3: Restless Leg Syndrome, Neck and Shoulder Pain
- Module 4: Risk Aversion
- Module 5: Internet Use
- Module 6: Loneliness, Stress, and Social Support/Burden
- Module 8: Elsa Health Questions
- Module 9: Numeracy
- Module 10: Positive Well-Being
- Module 11: Later Life Education
- Module 12: Subjective Uncertainty about Stock Market Returns

#### Rules - conditions

The conditions that allow a respondent to get a question or sequence of questions have been included in the codebook. However, the programming of the instrument actually reused blocks of programming for similar sequences (i.e the questions about people in the household, the pension questions, the follow up for disability programs). While these questions are similar, sometimes the conditions to get them or the pattern within the sequence itself are not the same. We have eliminated these discrepancies wherever possible but you will find some rules that should not apply to some sequences here and there. We strongly recommend that you also check the Box and Arrow questionnaire whenever you have a question about flow and who should be in a given sequence. The main difference is that the Box and Arrow will have the immediate rules for a skip, while the codebook lists all the conditions necessary for the respondent to be given a sequence of questions.

#### 8E. Explanation of the difference between tracker and 2002 final data

There are some areas where the tracker file and the 2002 final release data are different. Below is a list of these situations.

GSUBHH - The tracker file assigns a '9' for all new spouses in 2002 for the previous wave's subhh number. The 2002 final release data, however, will list the subhh number of the household that the new spouse married into, since the new spouse will 'inherit' many of the relationships from that household.

HPN\_SP - There will be a spouse pn in the ID variables for a respondent, even if that spouse did not give an interview in 2002, and they will be referenced in the household sections by their HPN\_SP. Also, for a non-original R, we assign a spouse pn even though we will never interview that person, and track the spouse's influence on the household with that number. The tracker file will not reference these spouses.

Family and Financial R assignments - As we keep a record of a non-interviewed spouse in the data, we also release their assigned role in the

household (i.e. family or financial r) with a listing for hpn\_fin, hpn\_fam in the household sections. The tracker file will simply list the assignments of the people interviewed. Therefore, for example, we would have a nfinal blank line in section H for a household where '020' was the non-interviewed financial R. The tracker file would show that hh not having a financial R.

#### Pension Level Files

The 2002 early release (1.0) included files at the pension level for sections K and L. This was done to expedite the release to the public. In the 2002 final release these data have been restructured to the R level and incorporated into their respective sections - J, K and L. We have retained the ask order of the questions as much as possible, though there has been extensive renaming of variables. As a final note, the pension level still remains in the Preload section.

#### Interviewer Observations

New for this release is a section called Interviewer Observations. These are a set of questions (actually asked in a separate Blaise application) that the interviewers answered about interviewing situations and the environment of the respondents. There is some missing data throughout this file as the interviewer was not required (in 2002) to complete this for any given completed interview. Where available, we have filled in the missing cases with earlier interview attempts where the interviewer had completed them. They describe the respondent's home, general mood, who else was there, whether they received assistance, some notes on the mode, etc.

#### 8F. 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 branchpoint was not changed.

If a change was made to an actual amount variable followed by a series of unfolding bracket questions, the answers to the unfolding bracket questions were retained.

#### 8G. Width changes in the data

It is important to pay close attention to the widths of the variables as the convention has changed in 2002. For example, if the data at HJ115\_1 is 6 digits then width will also be 6 and if the data at HJ115\_2 is only 5 digits then the width will be 5. So, not all variables within the same loop will have the same width. This is particularly notable when dealing with DK and RF codes:

HJ115\_1(AMOUNT OF BENEFITS-1), DK is 999998, RF is 999999  
HJ115\_2(AMOUNT OF BENEFITS-2), DK is 99998, RF is 99999  
HJ115\_3(AMOUNT OF BENEFITS-3), DK is 998, RF is 999

## 9. Obtaining the Data

### 9A. Registration and Downloading the Data

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

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

### 9B. Conditions of Use

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

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

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

### 9C. Publications Based on Data

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

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

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

## 10. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the 2002 HRS Core (Final, Version 2.0) 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.

### 10A. 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/>

10B. 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](mailto:hqsquest@isr.umich.edu)

Postal service:

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

FAX: (734) 647-1186

## Appendix

Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments

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

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

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

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

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

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

### A1. Married Couple Stays Married.

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

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

Time 1

Household records

HHID=012345 ASUBHH=0

Respondent records

HHID=012345 PN=010 ASUBHH=0

HHID=012345 PN=020 ASUBHH=0

Time 2

Household records

HHID=012345 CSUBHH=0

Respondent records

HHID=012345 PN=010 CSUBHH=0

HHID=012345 PN=020 CSUBHH=0

### A2. Couple Divorces.

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

By the time of the second cross-section, the couple has divorced. Both respondents retain the HHID of 023456, but one is assigned a SUBHH of 1 and the

---

<sup>1</sup> In addition, a SUBHH of 3 or 4 indicates the "household" of 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.

other is assigned a SUBHH of 2. Each original respondent retains his and her PN of 010 and 020, respectively.

Time 1

Household records

HHID=023456 ASUBHH=0

Respondent records

HHID=023456 PN=010 ASUBHH=0

HHID=023456 PN=020 ASUBHH=0

Time 2

Household records

HHID=023456 CSUBHH=1

HHID=023456 CSUBHH=2

Respondent records

HHID=023456 PN=010 CSUBHH=1

HHID=023456 PN=020 CSUBHH=2

A3. One or Both Respondents Die.

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

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

Time 1

Household records

HHID=034567 ASUBHH=0

Respondent records

HHID=034567 PN=010 ASUBHH=0

HHID=034567 PN=020 ASUBHH=0

Time 2

Household records

HHID=034567 CSUBHH=0

HHID=034567 CSUBHH=3 (in exit interview)

Respondent records

HHID=034567 PN=010 CSUBHH=3 (in exit interview)

HHID=034567 PN=020 CSUBHH=0

A4. Single Respondent Marries.

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

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

Time 1

Household records

HHID=045678 ASUBHH=0

Respondent records

HHID=045678 PN=010 ASUBHH=0

Time 2

Household records

HHID=045678 CSUBHH=0



Respondent records

HHID=045678 PN=010 CSUBHH=0  
HHID=045678 PN=011 CSUBHH=0

A5. Couple Divorces, One Respondent Remarries and Divorces.

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

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

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

Time 1

Household records

HHID=056789 ASUBHH=0

Respondent records

HHID=056789 PN=010 ASUBHH=0  
HHID=056789 PN=020 ASUBHH=0

Time 2

Household records

HHID=056789 CSUBHH=1  
HHID=056789 CSUBHH=2

Respondent records

HHID=056789 PN=010 CSUBHH=2  
HHID=056789 PN=011 CSUBHH=2  
HHID=056789 PN=020 CSUBHH=1

Time 3

Household records

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

Respondent records

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

A6. Couple Divorces and Marries Again.

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

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

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

Time 1  
Household records  
    HHID=067890 ASUBHH=0  
Respondent records  
    HHID=067890 PN=010 ASUBHH=0  
    HHID=067890 PN=020 ASUBHH=0

Time 2  
Household records  
    HHID=067890 CSUBHH=1  
    HHID=067890 CSUBHH=2  
Respondent records  
    HHID=067890 PN=010 CSUBHH=1  
    HHID=067890 PN=020 CSUBHH=2

Time 3  
Household records  
    HHID=067890 FSUBHH=7  
Respondent records  
    HHID=067890 PN=010 FSUBHH=7  
    HHID=067890 PN=020 FSUBHH=7

#### A7. Married Couple with Children and Siblings.

At the time of the first cross-section, sample household with a HHID of 078901 contains two respondents assigned PNs of 010 and 020, respectively. Associated with the household are three children with OPNs of 101, 102, and 103, and two siblings with OPNs of 051 and 052. All seven persons will keep those same PNs and OPNs across time. A friend who lives with the respondents is assigned an OPN of 080. The friend will keep her OPN of 080 across time only if she is a household member at each wave.

Time 1  
Household records  
    HHID=078901 ASUBHH=0  
Respondent records  
    HHID=078901 PN=010 ASUBHH=0  
    HHID=078901 PN=020 ASUBHH=0  
Household member/child records  
    HHID=078901 ASUBHH=0 OPN=101 (child)  
    HHID=078901 ASUBHH=0 OPN=102 (child)  
    HHID=078901 ASUBHH=0 OPN=103 (child)  
    HHID=078901 ASUBHH=0 OPN=051 (sibling)  
    HHID=078901 ASUBHH=0 OPN=052 (sibling)  
    HHID=078901 ASUBHH=0 OPN=080 (friend)

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

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

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

The two children appear both in their father's SUBHH 2 and also in their mother's SUBHH 1 with their respective OPNs, 101, and 102. The three siblings appear in the SUBHH of their respective siblings and maintain their respective OPNs. The ex-wife's brothers appear as part of her SUBHH 1 with their OPNs of 051 and 052, respectively. The ex-husband's sister appears as part of his SUBHH

2 with her OPN of 061. The friend has moved out and does not appear in either household.

Time 1

Household records

HHID=089012 ASUBHH=0

Respondent records

HHID=089012 PN=010 ASUBHH=0

HHID=089012 PN=020 ASUBHH=0

Household member/child records

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

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

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

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

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

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

Time 2

Household records

HHID=089012 CSUBHH=1

HHID=089012 CSUBHH=2

Respondent records

HHID=089012 PN=010 CSUBHH=2

HHID=089012 PN=020 CSUBHH=1

Household member/child records

HHID=089012 CSUBHH=1 OPN=101 (child)

HHID=089012 CSUBHH=1 OPN=102 (child)

HHID=089012 CSUBHH=1 OPN=051 (her brother)

HHID=089012 CSUBHH=1 OPN=052 (her brother)

HHID=089012 CSUBHH=2 OPN=101 (child)

HHID=089012 CSUBHH=2 OPN=102 (child)

HHID=089012 CSUBHH=2 OPN=061 (his sister)

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

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

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

Time 1

Household records

HHID=090123 ASUBHH=0

Respondent records

HHID=090123 PN=010 ASUBHH=0

HHID=090123 PN=020 ASUBHH=0

Time 2

Household records

HHID=090123 CSUBHH=1

HHID=090123 CSUBHH=2

Respondent records

HHID=090123 PN=010 CSUBHH=1

HHID=090123 PN=011 CSUBHH=1 (new wife)

HHID=090123 PN=020 CSUBHH=2

Household member/child records

HHID=090123 CSUBHH=1 OPN=151 (his stepchild)  
HHID=090123 CSUBHH=1 OPN=152 (his stepchild)  
HHID=090123 CSUBHH=2 OPN=151 (her mom)