

HEALTH AND RETIREMENT STUDY  
1992 Core  
Final, Version 2.0  
August 2004  
Data Description and Usage:  
What's New in Version 2.0

---

TABLE OF CONTENTS

|                                                                          |    |
|--------------------------------------------------------------------------|----|
| OVERVIEW OF VERSION 2.00 RELEASE .....                                   | 1  |
| DATA AND PROGRAM STATEMENT FILES.....                                    | 2  |
| Number of Records and Number of Variables in Data Files.....             | 2  |
| File Extensions.....                                                     | 3  |
| CHANGES TO VARIABLES.....                                                | 3  |
| Respondent-level and PARENTS Files Variable Changes.....                 | 4  |
| Household-level, KIDS, SIBLINGS, and HHLIST Files Variable Changes ..... | 4  |
| CHANGES TO DATA VALUES .....                                             | 5  |
| Data Alerts .....                                                        | 5  |
| PRIMARY AND SECONDARY IDENTIFIERS .....                                  | 6  |
| Respondent-level and PARENTS Files Identifiers.....                      | 6  |
| Household-level, KIDS, SIBLINGS, and HHLIST Files Identifiers.....       | 7  |
| CODEBOOK INFORMATION .....                                               | 7  |
| Codebook Files .....                                                     | 7  |
| Codeframes for Identifiers and Other Added Variables .....               | 9  |
| IF YOU NEED TO KNOW MORE .....                                           | 12 |
| HRS Internet Site .....                                                  | 12 |
| Contact Information .....                                                | 12 |
| PREVIOUS DOCUMENTATION.....                                              | 13 |
| Data Description and Usage, Version 1.0 (datadesc.txt) .....             | 14 |

---

## OVERVIEW OF VERSION 2.00 RELEASE

This document is intended to serve as an overview of what's new in the version 2.0 release of the 1992 HRS Core (Final) data. The primary change is the provision of standard identification variables in character (string), instead of numeric format, to facilitate merging with the tracker file and later years' data files. Additional changes are listed below.

- The combined variable, HHIDPN, has been dropped. Other identification variables and AVERSION have been added. Other than these changes, no variables have been deleted or added in the version 2.00 release.
- Data alerts that necessitated changes to the data have been incorporated. As a result, both the number of records in some files and the values of some variables, including identification variables, have changed.
- The data and associated program statement files for SAS, SPSS, and Stata are provided with standard extensions.
- The 1992 distribution files are provided as .zip files instead of .exe files.
- A listing of codebook files, the variables contained therein, and the corresponding data file is provided. The codebook files have not changed since 1995; the frequencies in the codebook files are based on data available at that time.
- The questionnaire files are now distributed in Adobe .PDF format.
- OSIRIS data descriptors and extract data descriptors are no longer provided.
- MSA flag files are now included along with other data and program statement files.
- For historical, contextual, purposes, the document Data Description and Usage, Version 1.0 (datadesc.txt) is included, unedited, at the end of this document.

## DATA AND PROGRAM STATEMENT FILES

### *Number of Records and Number of Variables in Data Files*

The table below lists the number of records in the version 2.0 file and number of variables in each of the version 1.01 and version 2.0 files.

- The number of variables changes because of the change in identification variables; see the [CHANGES TO VARIABLES](#) section below for details.
- The number of records changes because of the incorporation of data alerts that necessitated changes to the data; see the [CHANGES TO DATA VALUES](#) section below for details.

| File     | Number of Records V1.01 | Number of Records V2.0 | Records Deleted | Number of Variables V1.01 | Number of Variables V2.0 | Net number of Variables Added |
|----------|-------------------------|------------------------|-----------------|---------------------------|--------------------------|-------------------------------|
| PARENTS  | 25,444                  | 25,444                 | 0               | 46                        | 54                       | 8                             |
| KIDS     | 24,697                  | 24,697                 | 0               | 40                        | 48                       | 8                             |
| HHLIST   | 20,268                  | 20,268                 | 0               | 13                        | 21                       | 8                             |
| SIBLINGS | 19,000                  | 18,992                 | -8              | 26                        | 34                       | 8                             |
| EMPLOYER | 12,652                  | 12,652                 | 0               | 1,139                     | 1,142                    | 3                             |
| HEALTH   | 12,652                  | 12,652                 | 0               | 788                       | 791                      | 3                             |
| MSAFLAG  | 12,652                  | 12,652                 | 0               | 6                         | 9                        | 3                             |
| HOUSEHLD | 7,607                   | 7,607                  | 0               | 1,029                     | 1,032                    | 3                             |
| WIDOWED  | 5,045                   | 5,045                  | 0               | 48                        | 51                       | 3                             |
| MODULEG  | 1,124                   | 1,124                  | 0               | 23                        | 26                       | 3                             |
| MODULED  | 820                     | 820                    | 0               | 22                        | 25                       | 3                             |

| File    | Number of Records V1.01 | Number of Records V2.0 | Records Deleted | Number of Variables V1.01 | Number of Variables V2.0 | Net number of Variables Added |
|---------|-------------------------|------------------------|-----------------|---------------------------|--------------------------|-------------------------------|
| MODULEF | 788                     | 788                    | 0               | 28                        | 31                       | 3                             |
| MODULEJ | 763                     | 763                    | 0               | 49                        | 52                       | 3                             |
| MODULEE | 745                     | 745                    | 0               | 19                        | 22                       | 3                             |
| MODULEC | 698                     | 698                    | 0               | 25                        | 28                       | 3                             |
| MODULEB | 692                     | 692                    | 0               | 35                        | 38                       | 3                             |
| MODULEA | 662                     | 662                    | 0               | 27                        | 30                       | 3                             |
| MODULEH | 560                     | 560                    | 0               | 35                        | 38                       | 3                             |
| MODULEK | 229                     | 229                    | 0               | 19                        | 22                       | 3                             |

### ***File Extensions***

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 EMPLOYER.SAS go with the EMPLOYER.DA data file.

In the previous versions other, non-standard extensions were used. These are designated in the table below. Files with “include” statements for SAS and SPSS are no longer provided.

|                             | Version 1.01 | Version 2.0 |
|-----------------------------|--------------|-------------|
| SAS program statements      | .SAI         | .SAS        |
| SAS include statements      | .SAS         | ---         |
| SPSS program statements     | .SPI         | .SPS        |
| SPSS include statements     | .SPS         | ---         |
| Stata dictionary statements | .DCT         | .DCT        |
| Stata do statements         | .DO          | .DO         |

## **CHANGES TO VARIABLES**

A summary of variables dropped, added, renamed, or converted appears below. Please be aware that the codebook has not been updated to reflect these changes. See [Codeframes for Identifiers and Other Added Variables](#) section below for detailed variable information.

**Respondent-level and PARENTS Files Variable Changes**

| Data Files                                                                |                                                                           | Dropped | Converted to Character (String) | Added                                                                    |
|---------------------------------------------------------------------------|---------------------------------------------------------------------------|---------|---------------------------------|--------------------------------------------------------------------------|
| EMPLOYER<br>HEALTH<br>WIDOWED<br>MSAFLAG<br>MODULEA<br>MODULEB<br>MODULEC | MODULED<br>MODULEE<br>MODULEF<br>MODULEG<br>MODULEH<br>MODULEJ<br>MODULEK | HHIDPN  | HHID<br>PN<br>ASUBHH            | APN_SP<br>AFAMR<br>AFINR<br>AVERSION                                     |
| PARENTS                                                                   |                                                                           |         |                                 | HHID<br>PN<br>ATYPEPAR<br>ASUBHH<br>APN_SP<br>AFAMR<br>AFINR<br>AVERSION |

**Household-level, KIDS, SIBLINGS, and HHLIST Files Variable Changes**

| Data File        | Dropped | Converted to Character (String) | Renamed                                      | Added                                                                           |
|------------------|---------|---------------------------------|----------------------------------------------|---------------------------------------------------------------------------------|
| HOUSEHLD         | HHIDPN  | HHID<br>ASUBHH                  | Was: PN<br>Now: APN_FIN*                     | APN_FAM<br>APN_NFAM<br>APN_NFIN<br>AVERSION                                     |
| SIBLINGS<br>KIDS |         |                                 |                                              | HHID<br>ASUBHH<br>OPN<br>APN_FAM<br>APN_NFAM<br>APN_FIN<br>APN_NFIN<br>AVERSION |
| HHLIST           |         |                                 | Was: V2<br>Now: V1<br><br>Was: V1<br>Now: V2 | HHID<br>ASUBHH<br>OPN<br>APN_FAM<br>APN_NFAM<br>APN_FIN<br>APN_NFIN<br>AVERSION |

\*Converted to character (string)

## CHANGES TO DATA VALUES

### **Data Alerts**

Data alerts that necessitated changes to the data have been incorporated in version 2.0 release. Details are provided below.

#### **(1995-10-30) HEALTH: Calculated Age: V46**

This data alert contained a correction statement for V46 (R:AGE):

IF (V42 > V23) THEN V46 = V46-1

This correction has been applied to V46 in the version 2.0 release; a total of 4,891 cells were changed. However, we suggest you use BIRTHMO and BIRTHYR from the tracker file instead. The retired data alert may be found at our Web site on the Retired or Superseded Data Alerts page.

#### **(1996-02-14) EMPLOYER: Negative Job Durations: V3403, V3418, V3604, V3607**

Some values in the job duration variables were negative. Corrections fix at least some of the negative values (based on looking up the original interview) for V3403, V3418, V3604, and V3607. The corrections may be found at our Web site on the Retired or Superseded Data Alerts page. A total of 17 cells were changed for these variables.

#### **(1996-02-14) EMPLOYER: V3551**

Only 9 cases were originally coded as having valid data in question V3551 (G45:RATE: FND JB-FEW MOS); this was an error. The corrections may be found at our Web site on the Retired or Superseded Data Alerts page. A total of 617 cells were changed for this variable.

#### **(1996-02-14) EMPLOYER: V3537**

Only 9 cases were originally coded as having valid data in question V3537 (G42:RATE: FND JB-FEW MOS); this was an error. The corrections may be found at our Web site on the Retired or Superseded Data Alerts page. A total of 378 cells were changed for this variable.

#### **(1996-08-21) HEALTH: Missing Data Codes for V1538**

The missing data statements in SAS, SPSS, and Stata program statement files for V1538 (E40:\$TOT ED FIN ASST:IMP) in the HEALTH file were incorrect. Corrected statements are now provided that designate values greater than or equal to 9,999,996 as missing data. Formerly, values greater than or equal to 99,996 were designated as missing data. A total of 3,721 cases were affected.

#### **(1996-08-27) EMPLOYER: V2814 Top Code**

The missing data statements in SAS, SPSS, and Stata program statement files for V2814 (F22:NUM/EMP:ALL LOCATNS) in the EMPLOYER file were incorrect. Corrected statements are now provided that designate values greater than or equal to 999,979 as missing data. Formerly, values greater than or equal to 999,990 were designated as missing data. No cases were affected.

**(1998-11-24) HEALTH: Problematic Age Variable: V46**

This data alert described problems with V46 (R:AGE), but included no correction statements. We suggest you use BIRTHMO and BIRTHYR from the tracker file instead. The retired data alert may be found at our Web site on the Retired or Superseded Data Alerts page.

**(2003-02-25) HHLIST: Other Person Number Changes: V8302**

Changes were to be made to other person numbers, V8302 (PERSON NUMBER :HHL ) in the 1992 HRS Wave 1 HHLIST file. The corrections may be found at our Web site on the Retired or Superseded Data Alerts page. A total of 32 cells were changed for this variable.

**(2003-11-13) SIBLINGS: Other Person Number Changes: V8103**

As a result of corrections provided for the 1994 W2SIBS file, the changes needed to be made to person numbers in the 1992 SIBLINGS file to maintain cross-year consistency. The corrections may be found at our Web site on the Retired or Superseded Data Alerts page. A total of 93 cells were changed for this variable and eight cases were deleted.

**(2003-11-13) KIDS: Duplicates, Other Person Number Changes, and Other Changes: V8003 and Other Variables**

Two sets of changes were made to person numbers in the 1992 HRS Wave 1 KIDS file.

- One set was first issued 07/13/1995. Corrections provided for the 1994 W2KIDS file, necessitated changes to person numbers in the 1992 KIDS file to maintain cross-year consistency.
- Additional changes were issued 11/13/2003. A total of 76 duplicate combinations of Household Identifier and Person Number were present in the Wave 1 Public Release KIDS file. Code was provided to correct the problem. Changes were also made to V8004 through V8020 for a few cases.

The corrections may be found at our Web site on the Retired or Superseded Data Alerts page. A total of 163 cells were changed for V8003 variables and a total of 22 cells were changed for the other variables.

## **PRIMARY AND SECONDARY IDENTIFIERS**

Primary identifiers uniquely identify records in a file. Secondary identifiers are provided to facilitate merging with other files. The primary and secondary identifiers for each of the data sets are listed in the tables below.

***Respondent-level and PARENTS Files Identifiers***

| Data Files |         | Primary Identifiers | Secondary Identifiers              |
|------------|---------|---------------------|------------------------------------|
| EMPLOYER   | MODULED | HHID<br>PN          | ASUBHH<br>APN_SP<br>AFAMR<br>AFINR |
| HEALTH     | MODULEE |                     |                                    |
| WIDOWED    | MODULEF |                     |                                    |
| MSAFLAG    | MODULEG |                     |                                    |
|            | MODULEH |                     |                                    |
| MODULEA    | MODULEJ |                     |                                    |
| MODULEB    | MODULEK |                     |                                    |
| MODULEC    |         |                     |                                    |

| Data Files | Primary Identifiers    | Secondary Identifiers |
|------------|------------------------|-----------------------|
| PARENTS    | HHID<br>PN<br>ATYPEPAR |                       |

***Household-level, KIDS, SIBLINGS, and HHLIST Files Identifiers***

| Data Files | Primary Identifiers | Secondary Identifiers                      |
|------------|---------------------|--------------------------------------------|
| HOUSEHLD   | HHID<br>ASUBHH      | APN_FAM<br>APN_NFAM<br>APN_FIN<br>APN_NFIN |
| SIBLINGS   | HHID                |                                            |
| KIDS       | ASUBHH              |                                            |
| HHLIST     | OPN                 |                                            |

## CODEBOOK INFORMATION

### ***Codebook Files***

In 1992 codebook files were distributed in separate files for each section of the interview. Since there is not a one-to-one correspondence between codebook files and data files, the contents of the original codebook files and the data files in which the variables appear are noted in the table below. Variables with imputed versions (V1nnnn) appear in the codebook next to original variable (Vnnnn); only the original variable is listed in the table below.

For documentation of identification variables, see the [Codeframes for Identifiers and Other Added Variables](#) section of this document. For more recently calculated weights, see the current tracker file, available at our Web site. For more recently calculated imputations, see the imputation files, 1992 HRS Imputations (Final) (v.3.0), also available at our Web site.

| Codebook File | Contents                            | Level* | Variables Range In Codebook | Data Files                                                                                                                                |
|---------------|-------------------------------------|--------|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| 01_COVER.TXT  | Cover sheet and file description    | ---    |                             |                                                                                                                                           |
| 02_MCODE.TXT  | Master codes                        | ---    |                             |                                                                                                                                           |
| 03_CORE.TXT   | Identification and weight variables | ---    | V1-V15                      | EMPLOYER MODULED<br>HEALTH MODULEE<br>HHLIST MODULEF<br>HOUSEHLD MODULEG<br>KIDS MODULEH<br>MODULEA MODULEJ<br>MODULEB MODULEK<br>MODULEC |
| 04_0.TXT      | Facesheet and miscellany            | R      | V21-V144                    | HEALTH                                                                                                                                    |
| 05_A.TXT      | Section A: Demographic Background   | R      | V201-V249                   | HEALTH                                                                                                                                    |

| Codebook File              | Contents                                      | Level* | Variables<br>Range In<br>Codebook | Data Files |
|----------------------------|-----------------------------------------------|--------|-----------------------------------|------------|
| 06_B.TXT                   | Section B: Health Status                      | R      | V301-V542                         | HEALTH     |
| 07_D1.TXT                  | Section D: Housing                            | H      | V601-V744                         | HOUSEHLD   |
| 08_D2.TXT                  | Section D: Housing                            | H      | V801-V842                         | HOUSEHLD   |
| 09_E.TXT                   | Section E: Family Structure and Transfers     | R      | V901-V2622                        | HEALTH     |
| 10_F1.TXT                  | Section F: Employment of Respondent           | R      | V2701-V2905                       | EMPLOYER   |
| 11_F2.TXT                  | Section F: Employment of Respondent           | R      | V2904-V3201                       | EMPLOYER   |
| 12_F3.TXT                  | Section F: Employment of Respondent           | R      | V3202-V3349                       | EMPLOYER   |
| 13_G.TXT                   | Section G: Last Job (R not currently working) | R      | V3401-V3600                       | EMPLOYER   |
| 14_H1.TXT                  | Section H: Job History                        | R      | V3601-V3734                       | EMPLOYER   |
| 15_H2.TXT                  | Section H: Job History                        | R      | V3735-V3949                       | EMPLOYER   |
| 16_J1.TXT                  | Section J: Disability                         | R      | V4001-V4210                       | EMPLOYER   |
| 17_J2.TXT                  | Section J: Disability                         | R      | V4210-V4418                       | EMPLOYER   |
| 18_J3.TXT                  | Section J: Disability                         | R      | V4419-V4639                       | EMPLOYER   |
| 19_J4.TXT                  | Section J: Disability                         | R      | V4640-V4853                       | EMPLOYER   |
| 20_K.TXT                   | Section K: Retirement Plans                   | R      | V4901-V5032                       | EMPLOYER   |
| 21_L.TXT                   | Section L: Cognition                          | R      | V5101-V5127                       | HEALTH     |
| 22_M.TXT                   | Section M: Net Worth Other than Housing       | H      | V5201-V5362                       | HOUSEHLD   |
| 23_N1.TXT                  | Section N: Income                             | H      | V5401-V5729                       | HOUSEHLD   |
| 24_N2.TXT                  | Section N: Income                             | H      | V5730-V6206                       | HOUSEHLD   |
| 25_N3.TXT                  | Section N: Income                             | H      | V6207-V6536                       | HOUSEHLD   |
| 26_N4.TXT                  | Section N: Income                             | H      | V6550-V6566                       | HOUSEHLD   |
| 27_R.TXT                   | Section R: Health and Life Insurance          | H      | V6601-V6926                       | HOUSEHLD   |
| 28_S.TXT                   | Section S: Widowhood                          | R      | V7001-V7036                       | WIDOWED    |
| 29_X.TXT                   | Section X: Interviewer's Observations         | R      | V7101-V7144                       | HEALTH     |
| (see this document, below) | Urbanicity                                    | R      | V7504                             | MSAFLAG    |
| 30_KIDS.TXT                | Children information                          | K      | V8001-V8020                       | KIDS       |

| Codebook File | Contents                    | Level* | Variables<br>Range In<br>Codebook | Data Files                    |                               |
|---------------|-----------------------------|--------|-----------------------------------|-------------------------------|-------------------------------|
| 31_SIBS.TXT   | Sibling information         | S      | V8101-V8113                       | SIBLINGS                      |                               |
| 32_PARS.TXT   | Parent information          | P      | V8201-V8224                       | PARENTS                       |                               |
| 33_HHL.TXT    | Household listing           | M      | V8301-V8310                       | HHLIST                        |                               |
| 34_MOD1.TXT   | Experimental<br>Modules A-F | R      | V9001-V9084                       | MODULEA<br>MODULEB<br>MODULEC | MODULED<br>MODULEE<br>MODULEF |
| 35_MOD2.TXT   | Experimental<br>Modules G-K | R      | V9085-V9162                       | MODULEG<br>MODULEH            | MODULEJ<br>MODULEK            |

\* Where

R = respondent

H = household

C = child

S = sibling

P = parent

M = household member

### **Codeframes for Identifiers and Other Added Variables**

Since the changed or added identifiers are not documented in the original codebooks, the following information is provided.

### **Identification Variables**

#### **For Respondent-level Files and for PARENTS**

#### *Primary Identifiers*

```
=====
HHID      HOUSEHOLD IDENTIFIER
          Type: Character      Width: 6      Decimals: 0
.....
          000001-099587. Household Identifier range
=====

PN        PERSON NUMBER
          Type: Character      Width: 3      Decimals: 0
          .....
          010. Person Number
          020. Person Number
          030. Person Number
          040. Person Number
          Blank. INAP (Inapplicable) - No spouse interview (in PARENTS file)
=====
For PARENTS only:
ATYEPAR  1992 MOTHER OR FATHER PARENT INDICATOR
          Type: Character      Width: 1      Decimals: 0
          User note: This is a character version of the variable V8202
          .....
          1. Mother
          2. Father
=====
```

## Secondary Identifiers

```
=====
ASUBHH      1992 SUB-HOUSEHOLD IDENTIFIER
            Type: Character      Width: 1          Decimals: 0
            .....
            0. Original sample household - no split from divorce or
              separation of spouses or partners

=====

APN_SP      1992 SPOUSE/PARTNER PERSON NUMBER
            Type: Character      Width: 3          Decimals: 0
            .....
            010. Person Number
            020. Person Number
            030. Person Number
            040. Person Number
            Blank. INAP (Inapplicable) - No spouse interview

=====

AFAMR      1992 WHETHER FAMILY RESPONDENT
            Type: Numeric        Width: 1          Decimals: 0
            .....
            1. Yes
            4. No - household had no family respondents but the person
              number of this respondent appears as the APN_FAM in the household
              record.
            5. No
            Blank. INAP (Inapplicable) - No spouse interview (in PARENTS file)

=====

AFINR      1992 WHETHER FINANCIAL RESPONDENT
            Type: Numeric        Width: 1          Decimals: 0
            .....
            1. Yes
            4. No - household had no financial respondents but the person
              number of this respondent appears as the APN_FIN in the
              household record.
            5. No
            Blank. INAP (Inapplicable) - No spouse interview (in PARENTS file)

=====
```

## For Household-level Files and for KIDS, SIBLINGS, and HHLIST

### Primary Identifiers

```
=====
HHID        HOUSEHOLD IDENTIFIER
            Type: Character      Width: 6          Decimals: 0
            .....
            000001-099587. Household Identifier range

=====

ASUBHH      1992 SUB-HOUSEHOLD IDENTIFIER
            Type: Character      Width: 1          Decimals: 0
            .....
            0. Original sample household - no split from divorce or
              separation of spouses or partners

=====
For KIDS, SIBLINGS, and HHLIST only

OPN         OTHER PERSON NUMBER
            Type: Character      Width: 3          Decimals: 0

            User note: See PERSON NUMBER MASTER CODE.
            .....
            010. Respondent Person Number
            020. Respondent Person Number
            030. Respondent Person Number
            031-034. Other Person Number (R's Parents)
```

040. Respondent Person Number  
 041-044. Other Person Number (R's Parents)  
 051-062. Other Person Number (R's Grandparents)  
 101-399. Other Person Number (Children)  
 401-499. Other Person Number (Grandchildren)  
 501-699. Other Person Number (Siblings)  
 701-799. Other Person Number (Other Relatives)  
 901-998. Other Person Number (non-Relative)

## Secondary Identifiers

APN\_FAM 1992 FAMILY RESP PERSON NUMBER  
 Type: Character Width: 3 Decimals: 0  
 .....  
 010. Person Number  
 020. Person Number  
 030. Person Number  
 040. Person Number  
 Blank. No family respondent

APN\_NFAM 1992 NON-FAMILY RESP PERSON NUMBER  
 Type: Character Width: 3 Decimals: 0  
 .....  
 010. Person Number  
 020. Person Number  
 030. Person Number  
 040. Person Number  
 Blank. No spouse/partner; no non-family respondent

APN\_FIN 1992 FINANCIAL RESP PERSON NUMBER  
 Type: Character Width: 3 Decimals: 0  
 .....  
 010. Person Number  
 020. Person Number  
 030. Person Number  
 040. Person Number  
 Blank. No financial respondent

APN\_NFIN 1992 NON-FINANCIAL RESP PERSON NUMBER  
 Type: Character Width: 3 Decimals: 0  
 .....  
 010. Person Number  
 020. Person Number  
 030. Person Number  
 040. Person Number  
 Blank. No spouse/partner; no non-financial respondent

## Other Variables

### For All Files

AVERSION 1992 DATA RELEASE VERSION  
 Type: Numeric Width: 3 Decimals: 0  
 .....  
 3. Version 3, July 2004

## For MSAFLAG

```
=====
V7504   MSA/Non-MSA Flag (may be used as a rough indicator of urban/
        rural location of sample address)
        Type: Numeric           Width: 1           Decimals: 0
        .....
        0. Non-MSA County (rural)
        1. MSA County (urban)
=====
```

## IF YOU NEED TO KNOW MORE

This document is intended to serve as an overview of what's new in the version 2.0 release of the 1992 HRS Core (Final) 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.

### ***HRS Internet Site***

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site. Our URL is:

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

### ***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: [hrsrequest@isr.umich.edu](mailto:hrsrequest@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

## PREVIOUS DOCUMENTATION

The document Data Description and Usage, Version 1.0 (datadesc.txt) are included, unedited, in this document. **Much, but not all, of the material document is currently obsolete.** Please disregard the following information when it is in conflict with the proceeding information or with other, more recent, information at our Web site. As always, please contact us if you have additional questions.

***Data Description and Usage, Version 1.0 (datadesc.txt)***

---

HEALTH AND RETIREMENT STUDY  
WAVE I DATA

Public Release Version  
Data Description and Usage  
Version 1.0

May 08, 1995

---

Permission is not granted to reproduce any part of this document without the express written consent of the author(s) and/or staff of the Health and Retirement Study, University of Michigan.

---

This document was edited (and largely authored) by...

David Howell, HRS Staff  
Survey Research Center, Room 3253  
P. O. Box 1248  
Ann Arbor, MI 48106-1248

FAX: (313) 747-1186  
E-Mail Address: dahowell@umich.edu

...it is intended to serve as an outline and approach to using the data, not to be a comprehensive guide. I apologize for any errors found within.

---

This release of HRS Wave I data set is intended for use by the general public. By using the data set, you are agreeing to use conscientiously the data that has been freely provided to you, and that you will make no effort to identify the respondents therein. In addition, you are in good faith agreeing to send a copy of any publications you produce based on the data to the Health and Retirement Study at the address below.

The Health and Retirement Study  
Survey Research Center, 3200 Bay

## TABLE OF CONTENTS

### Sections

-----

...1 Acknowledgments

...2 Why this data structure and distribution method?

...3 Obtaining the Data

.....3-a Conditions of Use

.....3-b Internet Sites

...4 File descriptions

.....4-a Data Files

..... 4-a-1 Data Files Outline

.....4-b Codebook Files

.....4-b-1 Codebook Files Outline

.....4-c Dictionary Files

.....4-d Wave I Interview

.....4-d-1 Wave I Interview Outline

.....4-e Coding and Editing Instructions

...5 File usage

.....5-a Suggested Setup

.....5-b Decompressing the Files

..... 5-b-1 Using HRS1.BAT to Decompress the Files

..... 5-b-2 Decompressing the Files Yourself (or if  
you have problems with HRS1.BAT)

..... 5-b-3 Decompressed File Directories

...6 Data Description

.....6-a Imputations

.....6-b Masking for Confidentiality

...7 Some Useful Variables

...8 If You Have Special Needs or Problems

### Appendices

-----

...A Using OSIRIS with the Data

.....A-1. HRS Files for use with OSIRIS

...B Using SAS with the Data

.....B-1 HRS Files for use with SAS

.....B-2 Adding commands to the SAS programs (\*.SAS)

.....B-3 Submitting a SAS program

.....B-4 "Recalling" a SAS program

.....B-5 The SAS input statement file, and things you can do to it

.....B-6 Difficulties with the PROC CONVERT OSIRIS Engine (Dorothy Kempter, David Howell)

...C Using SPSS with the Data

.....C-1 HRS Files for use with SPSS

.....C-2 Setting your preferences (SPSS for Windows)

.....C-3 Creating an SPSS system file

.....C-4 Submitting a program (SPSS for Windows)

.....C-5 Adding commands to the SPSS program, without creating a system file

.....C-6 The SPSS input statements file, and things you can do to it

...D Concepts of Merging

.....D-1 Introduction

.....D-2 Family Descriptions

.....D-3 Merging Example

.....D-3-a Initial Files

.....D-3-b Individual-Level File Codebook

.....D-3-c Household-Level File Codebook

.....D-3-d Intermediate Step

.....D-3-e Final Step

.....D-3-f Codebook for Resulting "Family File"

.....D-4 Summary

...E Health and Retirement Study Staff

...F Glossary of Terms

...1 Acknowledgments

The Health and Retirement Study (HRS) is a national longitudinal study that focuses on persons born between the years 1931 and 1941 and their health, retirement, and economic status. It is "a cooperative agreement between the Institute for Social Research at

the University of Michigan and the National Institute on Aging."  
"Financial support for the project comes mainly from the National Institute on Aging, with supplementary funding provided by the Social Security Administration and the Assistant Secretary for Planning and Evaluation at the Department of Health and Human Services, and by the Pension and Welfare benefits Administration at the Department of Labor."

### ...2 Why this data structure and distribution method?

The data is currently stored in a series of files. All of these files are in ASCII format, with fixed-length fields, blocked at the end of each record. The primary goals of this storage method are:

- To allow generalized public distribution via the Internet
- To reduce the amount of storage space required by the data
- To create a data set that could be better accessed through a personal computer alone
- To allow a number of different types of data sets to be more easily created from the files provided, including individual, household, and longitudinal data sets
- To make merging and manipulating multiple types of files easier
- To allow users to exclude data of interest only to particular researchers from their working data set

All documentation files (save the interview itself) are in ASCII format to avoid compatibility problems. All documents are set up for 60 lines per page.

The data is being made available to the public via the Internet in hopes that a broad group of persons will make use of this very important collection of data.

### ...3 Obtaining the data

#### .....3-a Conditions of Use

HEALTH AND RETIREMENT STUDY ANNOUNCEMENT  
WAVE I PUBLIC RELEASE DATA

We, the Health and Retirement Study, request that persons who

download our data register with us before they do so. This allows us to distribute to our users information related to errors in the data, future releases, workshops, conferences, data distribution, and publication lists. The information you provide will not be for commercial use.

By downloading the data, you are agreeing to use conscientiously the data that have been freely provided to you, and that you will make no effort to identify the respondents therein. In addition, you are in good faith agreeing to send a copy of any publications you produce based on the data to the Health and Retirement Study at the address below.

|                                                               |  |
|---------------------------------------------------------------|--|
| +-----+                                                       |  |
| By submitting this information, I agree to the aforementioned |  |
| aforementioned conditions of use of the Health and Retirement |  |
| Study Wave I public release dataset.                          |  |
| +-----+                                                       |  |
| Name:                                                         |  |
| Date:                                                         |  |
| Electronic mail address:                                      |  |
| Postal address:                                               |  |
|                                                               |  |
|                                                               |  |
|                                                               |  |
| Phone number:                                                 |  |
| FAX number:                                                   |  |
| +-----+                                                       |  |

Please return a completed copy of the above information, either electronically to:

dahowell@isr.umich.edu

or by postal service to:

The Health and Retirement Study  
Survey Research Center, 3200 Bay  
P.O. Box 1248  
Ann Arbor, MI (USA) 48106-1248

Thank you for your assistance in this matter!

.....3-b Internet Sites

The Health and Retirement Study Wave I Public Release Data is

available by one of two methods, either the World-Wide-Web, or by FTP (File Transfer Protocol).

To access the data and other relevant information, you may point the World-Wide\_Web browser of your choice to the Health and Retirement Study Home Page at:

<http://www.umich.edu/~hrswww>

Alternately, you may choose to transfer the files by FTP by connecting to:

Site Name: ftp.psc.lsa.umich.edu  
User Name: anonymous  
...go to the directory /pub and download the document named HRS for further information

The Health and Retirement Study FTP sites are located on UNIX machines, courtesy of the Population Studies Center at the University of Michigan. Please use "anonymous" when prompted for your User Name, and your electronic mail address as a password.

#### ...4 File descriptions

##### .....4-a Data Files

The number of data files in the HRS Data Set (19 in all) may seem daunting at first. It eases the mind, however, to realize that there is no need to access every one of the files, but rather to determine by subject what is of interest and look at only the files containing relevant data. Most persons focus on their areas of interest and do not deal with the majority of the data available.

##### ..... 4-a-1 Data Files Outline

Each file contains a "core" set of variables which include the case identifier of both the respondent (Case ID) and their partner (Match ID), household identifier, and weights, as appropriate.

| File        | Contains                                                                                        |
|-------------|-------------------------------------------------------------------------------------------------|
| ----        | -----                                                                                           |
| EMPLOYED.DA | [Questions answered by all respondents;<br>one respondent per record]<br>Section F (Employment) |

Section G (Last Job)  
 Section H (Job History)  
 Section J (Disability)  
 Section K (Retirement Plans)

HEALTH.DA [Questions answered by most or all respondents;  
 one respondent per record]  
 Interview facesheet  
 Miscellaneous descriptive variables  
 Control file  
 Section A (Demographics and Background)  
 Section B (Health)  
 Section E (Family Structure)  
 Section L (Cognition)  
 Section X (Interviewer's Observations)

HHLIST.DA [One household member per record]  
 Information about household members

HOUSEHLD.DA [Questions answered by financially knowledgeable  
 respondents (R1s); one household per record]  
 Section D (Housing)  
 Section M (Net Worth other than Housing)  
 Section N (Income)  
 Section R (Health and Life Insurance)

KIDS.DA [One child per record]  
 Information about each child

MODULE?.DA [One respondent per line]  
 Experimental Modules

MODULEA.DA Module A (Physiological Measurements)  
 MODULEB.DA Module B (Activities of Daily Living (1))  
 MODULEC.DA Module C (Activities of Daily Living (2))  
 MODULED.DA Module D (Meta Memory)  
 MODULEE.DA Module E (Process Benefits)  
 MODULEF.DA Module F (Employment Alternatives)  
 MODULEG.DA Module G (Parent's Wealth)  
 MODULEH.DA Module H (Occupational Injuries)  
 MODULEJ.DA Module J (Health Risks)  
 MODULEK.DA Module K (Spending and Saving Preferences)

PARENTS.DA [One parent per record]  
 Information about each parent

SIBLINGS.DA [one sibling per line]

#### Information about each sibling

WIDOWED.DA [Questions answered by Secondary Respondents (R2s);  
one respondent per record]  
Section S (Widowhood)

#### ...4-b Codebook Files

The codebook files define what the codes associated with each variable number mean. The variable number is in the leftmost column. Also included are frequencies or means, as appropriate. Please note, however, that these statistics are UNWEIGHTED.

#### ..... 4-b-1 Codebook Files Outline

| File         | Pp.  | Description                                    |
|--------------|------|------------------------------------------------|
| ----         | ---- | -----                                          |
| 01_COVER.TXT | 2    | Cover sheet and file description               |
| 02_MCODE.TXT | 46   | Master codes                                   |
| 03_CORE.TXT  | 2    | Identification and weight variables            |
| 04_O.TXT     | 30   | Facesheet and miscellany                       |
| 05_A.TXT     | 26   | Section A (Demographic Background)             |
| 06_B.TXT     | 42   | Section B (Health Status)                      |
| 07_D1.TXT    | -    | -                                              |
| 08_D2.TXT    | 66   | Section D (Housing)                            |
| 09_E.TXT     | 40   | Section E (Family Structure and Transfers)     |
| 10_F1.TXT    | -    | -                                              |
| 11_F2.TXT    | -    | -                                              |
| 12_F3.TXT    | 114  | Section F (Employment of Respondent)           |
| 13_G.TXT     | 50   | Section G (Last Job (R not currently working)) |
| 14_H1.TXT    | -    | -                                              |
| 15_H2.TXT    | 78   | Section H (Job History)                        |
| 16_J1.TXT    | -    | -                                              |
| 17_J2.TXT    | -    | -                                              |
| 18_J3.TXT    | -    | -                                              |
| 19_J4.TXT    | 186  | Section J (Disability)                         |
| 20_K.TXT     | 26   | Section K (Retirement Plans)                   |
| 21_L.TXT     | 12   | Section L (Cognition)                          |
| 22_M.TXT     | 34   | Section M (Net Worth Other than Housing)       |
| 23_N1.TXT    | -    | -                                              |
| 24_N2.TXT    | -    | -                                              |
| 25_N3.TXT    | -    | -                                              |
| 26_N4.TXT    | 130  | Section N (Income)                             |
| 27_R.TXT     | 46   | Section R (Health and Life Insurance)          |
| 28_S.TXT     | 12   | Section S (Widowhood)                          |
| 29_X.TXT     | 18   | Section X (Interviewer's Observations)         |

|             |      |                          |
|-------------|------|--------------------------|
| 30_KIDS.TXT | 10   | Children information     |
| 31_SIBS.TXT | 6    | Sibling information      |
| 32_PARS.TXT | 12   | Parent information       |
| 33_HHL.TXT  | 4    | Household listing        |
| 34_MOD1.TXT | -    | -                        |
| 35_MOD2.TXT | 70   | Experimental Modules A-K |
| -----       | ---- |                          |
| TOTAL       | 1062 |                          |

#### .....4-c Dictionary Files

The dictionary files are those decompressed from the file OSIRIS.EXE (probably in the directory C:\HRS\WAVE1\OSIRIS). They are in ASCII/text format, and a file is present that corresponds to each of the 19 data files.

Files with the extension \*.DI are a Type 5 OSIRIS dictionary. Type 1 OSIRIS dictionaries can be obtained on a limited basis by special request.

Of more interest for documentation purposes are the files with the \*.DIL extension. These files are dictionary listings produced by OSIRIS from the Type 5 dictionaries. The pages are formatted and fairly easy to read. One variable is listed on each line, with the following information across the page:

- \* Variable number/name (VAR#)
- \* Variable label (VARIABLE NAME)
- \* First column at which the variable appears (COL)
- \* Width of the variable (WIDTH)
- \* Number of decimal points the variable has (NDEC)  
[for example, if NDEC=2, a data point might be 2.02;  
if NDEC=5, a data point might be 250.00001]
- \* Variable type (TYPE)  
[C is of character type, A is of alpha-numeric type]
- \* First missing data code (MDCODE1)  
[If present and less than MDCODE2, signifies that  
all codes equal to or less than this number are to  
be designated as missing data; if blank or equal to  
or greater than MDCODE2, ignore]
- \* Second missing data code (MDCODE2)  
[If present, all codes equal to or greater than this  
number are designated as missing data]

...note that there is additional information available in the listing, but these are the columns of most importance to you.

#### .....4-d Wave I Interview

A copy of the Wave I interview is available in the compressed file IVIEW.EXE. It is the only part of the Wave I Public Release documentation that is not in ASCII/text form. Each of the files listed in the next section is in WordPerfect 5.1 format, because that was how it was originally entered, and adequate conversion to text form was not feasible.

#### .....4-d-1 Wave I Interview Outline

| File         | Description                                |
|--------------|--------------------------------------------|
| ----         | -----                                      |
| HRS_PROD.A   | Section A (Demographic Background)         |
| HRS_PROD.B   | Section B (Health Status)                  |
| HRS_PROD.BRO | Study Fact Brochure                        |
| HRS_PROD.CVP | Primary Cover Sheet                        |
| HRS_PROD.CVS | Secondary Cover Sheet                      |
| HRS_PROD.D   | Section D (Housing)                        |
| HRS_PROD.E   | Section E (Family Structure and Transfers) |
| HRS_PROD.EF1 | Family Booklet Part I                      |
| HRS_PROD.EF2 | Family Booklet Part II                     |
| HRS_PROD.F   | Section F (Employment of Respondent)       |
| HRS_PROD.FS  | Face Sheet                                 |
| HRS_PROD.G   | Section G (Last Job)                       |
| HRS_PROD.H   | Section H (Job History)                    |
| HRS_PROD.IDC | Interviewer Data Card                      |
| HRS_PROD.IRC | Interviewer Reference Card                 |
| HRS_PROD.J   | Section J (Disability)                     |
| HRS_PROD.K   | Section K (Retirement Plans)               |
| HRS_PROD.L   | Section L (Cognition)                      |
| HRS_PROD.M   | Section M (Net Worth Other than Housing)   |
| HRS_PROD.N   | Section N (Income)                         |
| HRS_PROD.R   | Section R (Health and Life Insurance)      |
| HRS_PROD.RB  | Respondent Booklet                         |
| HRS_PROD.S   | Section S (Widowhood)                      |
| HRS_PROD.SCR | Call Record                                |
| HRS_PROD.X   | Section X (Interviewer's Observations)     |
| REL1MOD.A    | Module A (Physiological Measurements)      |
| REL1MOD.B    | Module B (Activities of Daily Living (1))  |
| REL1MOD.C    | Module C (Activities of Daily Living (2))  |
| REL2MOD.D    | Module D (Meta Memory)                     |
| REL2MOD.F    | Module F (Employment Alternatives)         |
| REL2MOD.G    | Module G (Parents' Wealth)                 |
| REL2MOD.H    | Module H (Occupational Injuries)           |

REL2MOD.J        Module J (Health Risks)  
REL3MOD.K        Module K (Spending and Saving Preferences)

#### .....4-e Coding and Editing Instructions

Wave I of HRS was a "paper-and-pencil" interview, and therefore required coding and editing by trained staff members. The instructions to those staff members are present in the compressed file INSTRUCT.EXE for those who are interested.

#### ...5 File usage

##### .....5-a Suggested Setup

While a particular setup is not required for using the HRS files, we do recommend the following.

Create the path C:\HRS\WAVE1 on your hard drive by typing:

```
md C:\HRS\WAVE1
```

Copy all of the files you retrieved from the HRS server to that directory. If you are in a directory with only those files in it, you may do so by typing:

```
copy *.* C:\HRS\WAVE1
```

By using this directory structure, you insure that HRS1.BAT will work appropriately with your files, organize your files well, and make our support staff's job much easier.

##### .....5-b Decompressing the Files

In DOS, run a listing of the files in the directory C:\HRS\WAVE1 (using the DIR command) to insure that the files you copied from the HRS FTP Server are in it.

At this point, you may choose to decompress the files by one of two methods: using HRS1.BAT, or doing it yourself. HRS1.BAT was written and tested under MSDOS 6.21, so it is somewhat possible that it may not run well under a previous version. You may type...

```
ver
```

...to determine what version of MSDOS your system is running on. If the program does not work (and the appropriate files are in

C:\HRS\WAVE1), it will be an obvious problem that you will notice. The decompressed files will either be there (it worked!) or they will not (it did not work!) once you are done.

If you plan to use HRS1.BAT, go now to Section 5-b-1 (the following one). If you are having trouble with HRS1.BAT and want to decompress the files yourself, turn to Section 5-b-2.

#### ..... 5-b-1 Using HRS1.BAT to Decompress the Files

Now you may run HRS1.BAT. Make sure you are in the directory C:\HRS\WAVE1 where both the compressed files and HRS1.BAT are present. Then type:

HRS1

You may type the .bat extension as well, but it is not required.

The first screen to appear should say that the program was written by myself (David Howell) and runs under MSDOS 6.21. It also reminds you that, as with all batch files, you may press the [CTRL] key and [BREAK] key at any time to halt the program. A brief reminder of responsible use of the data is included. You are then prompted to press a key to continue.

The second screen of HRS1.BAT reminds you that you need to have created the directory C:\HRS\WAVE1 on your hard drive. It also lists the compressed (self-extracting) files that you may decompress with HRS1.BAT and what they contain. If you have completed the instructions as stated, you may press 'Y' (yes) to go on. Pressing 'N' (for no) will result in the program halting execution so you may make the appropriate changes and run it again.

The third screen of HRS1.BAT is the one in which you choose which of the files you wish to decompress. The files, along with their compressed size, decompressed size, and number of decompressed files are in the table below. Make sure you have enough unused bytes on your hard drive to decompress the files you need. You can check this with the DIR command. When the computer prompts you to extract (decompress) a particular file, answer 'Y' (yes) if you wish to do so and 'N' (no) if you do not. You only need decompress whichever of the files that you want. For example, if you do not want the Wave I Interview, you need not decompress the associated file "IVIEW.EXE". You may not discriminate amongst files within a compressed file, however. If you only want the file PARENTS.DA when using HRS1.BAT, you need to

extract it and all of the other data files along with it.  
 (You may avoid this problem by using a variety of Shareware programs which allow you to UNZIP [decompress] only the files you want within the compressed file.)

| File         | Size       |              | Contents                        |
|--------------|------------|--------------|---------------------------------|
|              | Compressed | Decompressed |                                 |
| CODEBOOK.EXE | 293,793    | 2,065,885    | Codebook files                  |
| DATA.EXE     | 6,640,567  | 68,971,180   | Data files                      |
| INSTRUCT.EXE | 90,243     | 276,883      | Coding and editing instructions |
| OSIRIS.EXE   | 484,080    | 2,418,793    | OSIRIS-related files            |
| SAS.EXE      | 143,713    | 608,544      | SAS-related files               |
| SPSS.EXE     | 103,723    | 359,272      | SPSS-related files              |
| W1_IW.EXE    | 97,695     | 278,519      | Wave I interview                |

HRS1.BAT will print 'End of program.' to the screen and return you to the DOS prompt when it is done. Look in the appropriate directories (Section 5-b-3 to follow) for the decompressed files. The directory paths within the SAS and SPSS programs are already set so that they will work within this directory structure, assuming the necessary files are present.

..... 5-b-2 Decompressing the Files Yourself (or if you have problems with HRS1.BAT)

You may choose to decompress the files yourself, if you wish. Use the make directory (md) command in DOS to create the necessary directory paths to decompress the files in (see Table 4-4). Copy the compressed version of the file (the one with the .EXE extension) into the appropriate directory as shown in Section 5-b-3.

Also reference Section 5-b-3 to see what size the files are going to be after you decompress them. Insure that you have enough unused bytes on your disk drive to hold all of the decompressed files before decompressing. You may then enter the directories you created one at a time and run the executable files (they are set up to decompress themselves).

You will "see" the files being "inflated" into that directory. Try the DIR command once you are done to insure that they are there, and if they are, you are done!

Alternately, you may choose to place the HRS Wave I files into a

directory other than the C:\HRS\WAVE1 structure, but if you do so, be sure to alter the SAS and SPSS programs included to accommodate for the change. If you do not know how to do this, or have problems, either try using the directory structure given or obtaining help from your computing staff.

#### ..... 5-b-3 Decompressed File Directories

After decompression, whether you used HRS1.BAT or decompressed them yourself, the file directories should be as follows if you followed our recommendations:

| Path                  | Will contain                    |
|-----------------------|---------------------------------|
| ----                  | -----                           |
| C:\HRS\WAVE1\CODEBOOK | Codebook files                  |
| C:\HRS\WAVE1\DATA     | Data files                      |
| C:\HRS\WAVE1\INSTRUCT | Coding and editing instructions |
| C:\HRS\WAVE1\OSIRIS   | OSIRIS files                    |
| C:\HRS\WAVE1\SAS      | SAS files                       |
| C:\HRS\WAVE1\SPSS     | SPSS files                      |
| C:\HRS\WAVE1\W1_IW    | Wave I Interview                |

#### ...6 Data Description

Respondent Type. In all, 12,652 respondents were successfully interviewed and coded for HRS Wave I. Respondents were chosen prior to the interview to be either "primary" or "secondary" respondents.

If a person was single, they were designated the primary respondent (R1). In a married pair or partnership, the person most knowledgeable of household financial matters was designated the primary respondent. Primary respondents were to answer questions regarding housing, assets, and liabilities.

Secondary respondents (R2s) are defined as the respondent who is not the primary respondent in a married pair or partnership. In a household with a single respondent, there is no secondary respondent.

It should be noted that whether a respondent is the primary or secondary respondent is really of no value analytically other than to methodologists. It is purely a designation designed so that the respondent most knowledgeable about the household's finances would answer the appropriate questions.

You may determine if a particular respondent is a single R1, paired R1, or paired R2 by their case identifier (variable one [V1] of the dataset). If the Case ID is in the range 1-1999, the respondent is a single R1. A Case ID in the range 10000-19999, reveals that the respondent is a paired R1. If the Case ID is in the range 20000-29999, the respondent is a paired R2.

As shown in Table 6-1, there were 2,373 single R1s and 5,234 married/partnered (paired) R1s (7,607 primary respondents in all). In addition, there were 5,045 R2s (all married or partnered).

Table 6-1. Respondent Types

| Respondent Type | N      | Respondent Type | N      |
|-----------------|--------|-----------------|--------|
| Single Rs       | 2,373  | R1s             | 7,607  |
| Paired R1s      | 5,234  |                 |        |
| Paired R2s      | 5,045  | R2s             | 5,045  |
| TOTAL           | 12,652 | TOTAL           | 12,652 |

Section E Respondent Status. Only one person per household was to answer questions concerning household and family members (Section E). Those persons who answered Section E were referred to as "Section E respondents". There was to be only one Section E respondent per household. In single households, the lone respondent was always the Section E respondent. In married/partnered (paired) households, the Section E respondent was chosen at question E1, variable 901 [V901]. Females, males answering for females who refused the interview, and primary respondents in a same-sex couple (codes 2-4 in V901) were chosen to be Section E respondents. All others were not to answer Section E (codes 1 and 5 in V901).

Table 6-2 of the following page shows the number of respondents, by respondent type, which answered Section E. There were 7,547 respondents who answered Section E, and 5,105 who did not. Of the 7,547 respondents who answered Section E, 4,402 were primary respondents (2,373 singles and 2,029 paired R1s) and 3,145 were secondary respondents.

It is important to note that the total number of Section E respondents (7,547) is not equal to the number of households (7,702). There is no Section E respondent for 55 households.

Table 6-3. Respondent Types by Section E Status

| Respondent Type | Answered Section E? |       | R Type | Answered Section E? |       |
|-----------------|---------------------|-------|--------|---------------------|-------|
|                 | Yes                 | No    |        | Yes                 | No    |
| Single Rs       | 2,373               | 0     | R1s    | 4,402               | 3,205 |
| Paired R1s      | 2,029               | 3,205 |        |                     |       |
| Paired R2s      | 3,145               | 1,900 | R2s    | 3,145               | 1,900 |
| TOTAL           | 7,547               | 5,105 | TOTAL  | 7,547               | 5,105 |

Households. The 12,652 respondents interviewed were from a total of 7,702 households. The breakdown, by respondent type, is found in Table 7-3, below. There were 2,373 households with a single respondent, and 5,329 with respondents who were paired (married/partnered). This might lead one to believe that there should be 13,031 respondents in all ( $2373 + (2 \times 5329)$ ), but in some cases, one of the respondents (which could be either the R1 or the R2) in the pair refused to do the interview. This is elaborated on in the text after Table 6-3.

Table 6-3. Respondent Types by Households

| Respondent Types | Households |
|------------------|------------|
| Single Rs        | 2,373      |
| Pair of Rs       | 5,329      |
| TOTAL            | 7,702      |

Refusals. As in any survey, refusals had to be dealt with. In the HRS, we chose to not include a record for those respondents who refused to answer the interview. Thus, every one of the 12,652 respondents who answered the interview have a record, but for those who did not answer the interview there are no records.

Single persons who did not answer the interview were left out of the data, and thus their household was left out of the data, too. They were the only person in the household to interview, so we had no information. In a household with two paired respondents,

however, if one respondent refuses, we still have information from the other respondent. The respondent who gave the interview would have a record in the dataset, and their spouse/partner would not. Thus, we have some cases where a valid respondent will not have a valid spouse/partner record. When a respondent has a spouse/partner who refused an interview, they will have a number in the range 30000-39999 as their Match ID.

As shown in Table 6-4 below, there were 379 paired (married/partnered) households in which one of the respondents refused the interview. Paired R1s refusing the interview accounted for 284 of the households, and paired R2s the other 95 households. There were 7,323 households in which all respondents did the interview (2,373 single households and 4,950 paired households). That makes for 7,702 households in all (7,323+379).

Table 6-4. Household Types

| Household Types  | Respondents | Respondents | Households |
|------------------|-------------|-------------|------------|
| Single R, no R1  | 2,373       | 2,373       | 2,373      |
| R1 with valid R2 | 4,950       | 9,900       | 4,950      |
| R2 with valid R1 | 4,950       |             |            |
| R1, no R2        | 284         | 379         | 379        |
| R2, no R1        | 95          |             |            |
| TOTAL            | 12,652      | 12,652      | 7,702      |

Variables and Records Per File. Table 6-5 on the page to follow lists the number of variables and records located in each file in the HRS Wave I (ICPSR Version) data set. These statistics can be used in a variety of ways.

First of all, they may be used to check that your data is correct as you received it. There is always the possibility (though presumably small) of a transfer, extraction, or conversion problem when receiving data. If the data you have match the statistics in Table 6-5, the chance that your data has been corrupted is less.

From the amount of records, you can tell how many persons or households (as appropriate) are in each file.

Table 6-5. Number of variables and Records Per File

| File        | Variables | Records | Sorted By |
|-------------|-----------|---------|-----------|
| EMPLOYER.DA | 1,136     | 12,652  | V1 (Case) |
| HEALTH.DA   | 784       | 12,652  | V1 (Case) |
| HHLIST.DA   | 13        | 20,268  | V2 (HH)   |
| HOUSEHLD.DA | 1,025     | 7,607   | V2 (HH)   |
| KIDS.DA     | 40        | 24,697  | V2 (HH)   |
| MODULEA.DA  | 23        | 662     | V1 (Case) |
| MODULEB.DA  | 31        | 692     | V1 (Case) |
| MODULEC.DA  | 21        | 698     | V1 (Case) |
| MODULED.DA  | 18        | 820     | V1 (Case) |
| MODULEE.DA  | 15        | 745     | V1 (Case) |
| MODULEF.DA  | 24        | 788     | V1 (Case) |
| MODULEG.DA  | 19        | 1,124   | V1 (Case) |
| MODULEH.DA  | 31        | 560     | V1 (Case) |
| MODULEJ.DA  | 45        | 763     | V1 (Case) |
| MODULEK.DA  | 15        | 229     | V1 (Case) |
| PARENTS.DA  | 46        | 25,444  | V2 (HH)   |
| SIBLINGS.DA | 26        | 19,000  | V2 (HH)   |
| WIDOWED.DA  | 45        | 5,045   | V1 (Case) |

For individual-level files (such as EMPLOYER.DA, HEALTH.DA, WIDOWED.DA, and the modules files), the number of records is equal to the number of respondents in the file. This is because there is a single respondent/record per line. For example, there were 820 persons who answered Module D (MODULED.DA), 7,607 primary respondents (HOUSEHLD.DA), and 12,652 respondents in all (HEALTH.DA).

For household-level files (such as HHLIST.DA, KIDS.DA, PARENTS.DA, and SIBLINGS.DA), the number of records equal the number of persons in the file. There were 20,268 persons total in 7,702 households represented (from HHLIST.DA). Children totalled 24,697 in number. A total of 25,444 parents were asked about, and 19,000 siblings were recorded.

#### .....6-a Imputations

A large number of variables were imputed in the Wave I public release data set. Those analysts who wish to use the imputations need do nothing out of the ordinary. For those who do not care

to use the imputed values, or perhaps just wish to know what the original value was, we have created imputation indicators.

Imputed variables are designated in two places, the codebook and in the variable label. Variables which are imputed have an associated variable that is its imputation indicator. The imputation indicator is always the variable number plus an additional 10,000. For example, if V100 were imputed, its imputation indicator would be V10100; for V9999 the imputation indicator would be V19999, etc.

The imputation indicators are one digit codes reflecting the original value of the data. The values of the imputation indicators are shown in the table below.

| Indicator<br>Value | Meaning (original value of variable)                                                                                                                                |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| -----              | -----                                                                                                                                                               |
| 1                  | Partially missing data: brackets were used to obtain the value                                                                                                      |
| 2                  | Original value was termed Inap.                                                                                                                                     |
| 3                  | Original value was not missing, and was inside the valid range of codes (most probably, a prior variable was imputed and changed the skip pattern of this variable) |
| 4                  | Original value was not missing, but was outside the valid range of codes                                                                                            |
| 5                  | Missing data: refused                                                                                                                                               |
| 6                  | Partially missing data: the range card was used to obtain the value                                                                                                 |
| 7                  | Missing data: loss/negative, DK/NA; ]<br>"Q not relevant to R"                                                                                                      |
| 8                  | Missing data: DK/don't know                                                                                                                                         |
| 9                  | Missing data: NA/Not ascertained                                                                                                                                    |

#### .....6-b Masking for Confidentiality

We at the Health and Retirement Study are dedicated to maintaining the confidentiality of our respondents. For that reason, a number of variables have been recoded or set to zero for the public release data set. The record of those changes follows.

Sample identifiers were removed.

In Section 0, V43 and V63, the respondent and partner's day of birth, was set to zero. The respondents' state of residence was

recoded to U.S. Census Division in V52 and V53.

The highest educational degree earned at V211 (A3d) was further grouped together to increase cell sizes.

Industry and occupation codes were recoded into a group of thirteen and seventeen codes, respectively, from the original three-digit U.S. Census code. The codes themselves may be found as part of the codebook. The variables recoded were:

| Section   | Industry | Occupation |
|-----------|----------|------------|
| -----     | -----    | -----      |
| Section F | V2719    | V2720      |
|           | V2820    | V2821      |
| Section G | V3406    | V3407      |
|           | V3431    | V3432      |
| Section H | V3605    | V3606      |
| Section J | V4303    | V4302      |
|           | V4403    | V4402      |
|           | V4438    | V4437      |
|           | V4506    | V4505      |

Finally, the parents' current state of residence was present in V9094 and V9095 for those respondents who were administered Module G. Those variables were also recoded down to U.S. Census Division level.

#### ...7 Some Useful Variables

In file HEALTH...

-----

|         |                                                                                                          |
|---------|----------------------------------------------------------------------------------------------------------|
| V1      | Case Identifier (Case ID)<br>- identifies individual respondents                                         |
| V2      | Household Identifier (Case ID)<br>- identifies individual households                                     |
| V3      | Match Identifier (Match ID)<br>- gives the case identifier for that respondent's partner, if appropriate |
| V11-V15 | Weights<br>- should be used with many, but not all, analyses                                             |
| V32     | Proxy Interview Flag                                                                                     |

- identifies cases where a proxy was used

V42,V44    Month and day of birth, respectively, of respondent

V46        Calculated age of respondent

V48,V49    Race/ethnicity of respondent

V207       Years of education of respondent

V225       Marital status of respondent

V901       Section E respondent status

In file EMPLOYER...

-----

V5362      Total net worth for household

V6559      Total household income

#### ...8 If You Have Special Needs or Problems

If you have any special requests or needs, feel free to contact the HRS Staff (see Appendix E). We will do our best to help. Suggestions and/or questions concerning data content, codes, methodology, and research-related topics are particularly welcome. Technical support beyond problems with the files themselves is limited.

Good luck!

#### APPENDIX A: USING OSIRIS WITH THE DATA (TECHNICAL NOTES)

##### ...A-1 HRS Files for use with OSIRIS

| Directory:           | Files: |
|----------------------|--------|
| C:\HRS\WAVE1\OSIRIS\ | *.DI   |
| C:\HRS\WAVE1\DATA\   | *.DA   |

Files with the suffix "DI" are OSIRIS Type 5 dictionaries. Use these files by substituting the file name into the DICTIN parameter of the input assignments.

Files with the suffix "DA" contain the data. Use these

files by substituting the file name the DATAIN parameter of the input assignments.

## APPENDIX B: USING SAS WITH THE DATA (TECHNICAL NOTES)

### ...B-1 HRS Files for use with SAS

| Directory:         | Files: |
|--------------------|--------|
| C:\HRS\WAVE1\SAS\  | *.SAS  |
| C:\HRS\WAVE1\SAS\  | *.SAI  |
| C:\HRS\WAVE1\DATA\ | *.DA   |

Files with the suffix "SAS" are SAS programs (two lines each) which you may add your own commands to the end of and execute. The only function of the two lines is to include the SAS input statements into your program for you.

Files with the suffix "SAI" are the SAS input statements.

Files with the suffix "DA" contain the data.

### ...B-2 Adding commands to the SAS programs (\*.SAS)

You may append any commands that you like to the files with the ".SAS" extension. After adding your commands to the end of the file, you may submit it as you would any SAS program. The SAS pseudo-file name ("file=<FILENAME>") is the same as the name of the file the data is coming from, without an extension.

Example:

1. An example program EX.SAS is...

```
%include 'c:\hrs\wave1\ex.sai';  
run;
```

2. You may then add your own commands to it, as shown.

```
%include 'c:\hrs\wave1\ex.sai';  
run;  
proc means file=ex;  
run;
```

3. At this point, "submit" your program (as explained in the next section). The output should contain the count,

mean, standard deviation, minimum, and maximum of every variable in the file IDV.DA.

...B-3 Submitting a SAS program (SAS for Windows)

1. Enter SAS for Windows.
2. Select "File" from the menu bar.
3. Select to add a program to the program editor.
4. Select the program with the .SAS extension which you wish to run. (It should be in C:\HRS\WAVE1\SAS\...)
5. Move the cursor to the program editor window.
6. Press the right mouse button to pull up a short menu.
7. Select "Locals".
8. Select "Submit".

...B-4 "Recalling" a SAS program (SAS for Windows)

Even though the program editor window goes blank after you run a SAS program, it is still in memory. Bring up a menu by pressing the right mouse key while within the program editor window. Select "Locals" and then "Recall". It will appear in the program editor window.

...B-5 The SAS input statements file, and things you can do to it

The SAS input statements for the data set EX (an example file) are as follows...

```
DATA ex;  
INFILE 'C:\HRS\WAVE1\DATA\ex.da' LRECL=15;
```

```
INPUT  
  V1 1-5  
  V2 6-10  
  V3 11-15  
  ;
```

```
IF V1=99999 THEN V1=.;  
IF V2=0 THEN V2=.;  
IF V2 GE 98 THEN V2=.;  
IF V3=00000 THEN V3=.;
```

```
LABEL  
  V1 ="IDENTIFIER FOR EXAMPLE FILE      "  
  V2 ="FIRST VARIABLE                   "  
  V3 ="SECOND VARIABLE                  "
```

;

Subsetting variables. First, let us say that we are only going to work with the variables V1 and V3. We can then delete all the lines associated with V2 (although this is not required). The resulting file will look as follows, and will not require as many system resources as the original file...

```
DATA ex;
  INFILE 'C:\HRS\WAVE1\DATA\ex.da' LRECL=15;

  INPUT
    V1 1-5
    V3 11-15
  ;

  IF V1=99999 THEN V1=.;
  IF V3=00000 THEN V3=.;

  LABEL
    V1 ="IDENTIFIER FOR EXAMPLE FILE"
    V3 ="SECOND VARIABLE"
  ;
```

The DATA statement. The first line of the file, the DATA statement, gives the data a pseudo-name, here "EX". From then on in the program, you refer to the data as EX. You can change the name EX to almost any name of your choosing. For example, you could change the pseudo-name to DAVID by editing the first line to read...

```
DATA david;
```

INFILE, LRECL. The second line of the file tells SAS where the data file is, and how long each record is. From EX.SAI, we know that SAS will look for the data file EX.DA in the directory C:\HRS\WAVE1\DATA\, and that each record in the file is 15 characters long. Do not alter the record length (LRECL) unless you change the data itself. You may, however, change the directory path. For example, let's say you wanted to store all of your data in the directory C:\DAVID\. After moving the file EX.DA to C:\DAVID\, you should change the second line of the file as follows...

```
INFILE 'C:\DAVID\ex.da' LRECL=15;
```

...now SAS will know where to look for the file.

Variable location definitions. The lines between the word INPUT and the next semicolon should not be changed (though they can be deleted if you are not going to use those variables). These lines define which columns in the data file correspond to the variables you are using.

Missing data values. All of the lines shown beginning with the word "IF" are missing data value statements. For example, "IF V10 GE 99999 THEN V10=." means that SAS should reassign all values of V10 greater than or equal to 99999 as missing data (".").

There is no rule that says you have to have missing data value statements at all. If you want a particular variable to have no missing/deleted values, remove the corresponding missing data value statements from the file.

Variable labels. The lines between the word LABEL and the following semicolon are the variable label statements. If you want a variable to have a different label, you can change it. If you do not want any labels at all, delete the entire block, resulting in...

```
DATA ex;
  INFILE 'C:\HRS\WAVE1\DATA\ex.da' LRECL=15;

  INPUT
    V1 1-5
    V3 11-15
    ;

  IF V1=99999 THEN V1=.;
  IF V3=00000 THEN V3=.;
```

#### ...B-6 Difficulties with the PROC CONVERT OSIRIS Engine (by Dorothy Kempter and David Howell)

If you plan on converting the OSIRIS dataset and dictionary provided into a SAS dataset using the SAS PROC CONVERT module, you should be aware of a difficulty that was discovered by prior users (Kathy Kitiyama and Beth Soldo).

The problem lies with how PROC CONVERT stores values on some computer platforms. On some platforms, the word length for the

PROC CONVERT module is only four bytes. Unfortunately, a word length of four bytes can not deal well with codes greater than 2,097,152. Thus, these numbers do not translate correctly. As you know, many of our variables have codes greater than this value. Specifically, any variable with a width of 7 in the OSIRIS dictionary (this includes virtually all 'dollar amount' variables) have range card and missing data codes in the 9-million range that will be affected.

This problem affects users running PROC CONVERT using the OSIRIS engine on AXP/VMS or UNIX installations or on PC-WINDOWS or SAS under OS2 -- i.e. any system using the IEEE Standard.

Releases running on VAX/VMS computers (v6.06, v6.07, and v6.08) or IBM mainframes do NOT produce the erroneous values. Computers may, however, be networked together in such a way that it is transparent to the user which system SAS is running on. The user may log into a VAX/VMS system, but the SAS system may not be running in that environment.

If you are using PROC CONVERT and you have any doubts about whether your system is storing these large numbers accurately, we strongly suggest that you check the upper end of your frequencies on E40 (V1538) against the table on the following page. If your values match those given, your computer platform is not one of those with the problem.

If your frequencies do not match the table, call the SAS Institute technical support line (919-677-8008) and they will help you work around the problem. Mention that you are running PROC CONVERT using the OSIRIS engine and refer to "USAGE NOTE U9257". You will need to re-pull your data.

We have been told that SAS is taking steps to revise the PROC CONVERT module so that this problem will be avoided in future releases. We apologize for any confusion this may have caused.

#### APPENDIX C: USING SPSS WITH THE DATA (TECHNICAL NOTES)

##### ...C-1 HRS Files for use with SPSS

|                    |        |
|--------------------|--------|
| Directory:         | Files: |
| C:\HRS\WAVE1\SPSS\ | *.SPS  |
| C:\HRS\WAVE1\SPSS\ | *.SPI  |

Files with the suffix "SPS" are SPSS programs (two lines each) which you may add your own commands to and execute. The only function of the two lines is to include the SPSS input statements into your program for you.

Files with the suffix "SPI" are the SPSS input statements.

Files with the suffix "DA" are the data files.

#### ...C-2 Setting your preferences (SPSS for Windows)

(Do this before you do anything else. When you exit SPSS for Windows, the preferences should be saved on your machine.)

1. Enter SPSS for Windows.
2. Select "Edit" from the menu bar.
3. Select "Preferences.." from the bottom of the menu.
4. Select "Output..." from the bottom right-hand corner of the window that appears.
5. Insure that the three boxes in the upper left-hand corner of the window that appears have "X"s in them. The boxes "Commands", "Error and Warnings", and "Resource Message" should all be marked with an "X". This tells SPSS for Windows to show you what the program is doing as it runs, not just tell you when there are errors.

#### ...C-3 Creating an SPSS system file

You may now convert any of the data sets into SPSS system files. Submitting any of the .SPS programs provided, as is, will create a file with the extension .SYS in the C:\HRS\WAVE1\SPSS directory on your computer. Load it in and use it as you would any SPSS system file. (For information on how to submit a program, read the next section.)

#### ...C-4 Submitting a program (SPSS for Windows)

(This option may be used to create a SPSS system file from the files provided.)

1. Enter SPSS for Windows.
2. Select "File" from the menu bar.
3. Select "Open" and then "SPSS Syntax".
4. Select the program with the .SPS extension which you

- wish to run. (It should be in C:\HRS\WAVE1\SPSS)
5. Do one of the following:
    - a. Place the cursor in the upper left hand corner of the window where your program is listed.
    - b. Select the entire program by using the mouse (if selected correctly, it will be highlighted in black.)...this step is because SPSS for Windows will only execute commands after the point where the cursor is located, or parts that are highlighted.
  6. Press the "Run" key in the window where your program is located.
- ...C-5 Adding commands to the SPSS program, without creating a system file

If you would like to add commands to the SPSS program file, but not create a system file, delete the last line of the program (it begins with "SAVE /OUTFILE"). After you delete the last line, add your own SPSS commands at the end.

Example:

1. The original program MODULEG.SPS was...

```
INCLUDE 'C:\HRS\WAVE1\SPSS\moduleg.spi'.  
SAVE /OUTFILE 'C:\HRS\WAVE1\SPSS\moduleg.sys'.
```

2. First you must delete the last line, resulting in...

```
INCLUDE 'C:\HRS\WAVE1\SPSS\moduleg.spi'.
```

3. You may then add your own commands to it, in the area shown...

```
INCLUDE 'C:\HRS\WAVE1\SPSS\moduleg.spi'.  
<PUT YOUR COMMANDS HERE>
```

4. Now submit the program as you learned previously to execute your commands.

...C-6 The SPSS input statements file, and things you can do to it

The SPSS input statements for the example data set EX are as follows...

```

FILE HANDLE ex /name='C:\HRS\WAVE1\DATA\ex.da' LRECL=15.

DATA LIST FILE=hhlist /
  V1 1-5
  V2 6-10
  V3 11-15
.

RECODE V1 (99999=SYSMIS).
RECODE V2 (0=SYSMIS).
RECODE V2 (98 THRU HIGHEST=SYSMIS).
RECODE V3 (0=SYSMIS).

VARIABLE LABELS
  V1 "IDENTIFIER FOR EXAMPLE FILE"
  V2 "FIRST VARIABLE"
  V3 "SECOND VARIABLE"
.
execute.

```

Subsetting variables. First, let us say that we are only going to work with the variables V1 and V3. We can then delete all the lines associated with V2 (although this is not required). The resulting file will look as follows, and will not require as many system resources as the original file...

```

FILE HANDLE ex /name='C:\HRS\WAVE1\DATA\ex.da' LRECL=15.

DATA LIST FILE=hhlist /
  V1 1-5
  V3 11-15
.

RECODE V1 (99999=SYSMIS).
RECODE V3 (0=SYSMIS).

VARIABLE LABELS
  V1 "IDENTIFIER FOR EXAMPLE FILE"
  V3 "SECOND VARIABLE"
.
execute.

```

File location. The first line of the file tells SPSS where the data file is, and how long each record is. From EX.SPI, we know that SPSS will look for the data file EX.DA in the directory C:\HRS\WAVE1\SPSS\. You may change the directory path.

For example, let's say you wanted to store all of your data in the directory C:\DAVID\. After moving the file EX.DA to C:\DAVID\, you should change the first line of the file as follows...

```
FILE HANDLE ex /name='C:\DAVID\ex.da' LRECL=15.
```

...now SPSS will know where to look for the file.

Variable location definitions. Statements starting with the second line and going through the first period should not be changed (though they can be deleted if you are not going to use those variables). These lines define which columns in the data file correspond to the variables you are using. Missing data values. All of the lines shown beginning with the word "RECODE" are missing data value statements. For example, "RECODE V10 (99999 THRU HIGHEST=SYSMIS)" means that SPSS should reassign all values of V10 greater than or equal to 99999 as missing data.

There is no rule that says you have to have missing data value statements at all. If you want a particular variable to have no missing/deleted values, remove the corresponding missing data value statements from the file.

Variable labels. The lines between the line "VARIABLE LABELS" and the following "EXECUTE" line are the variable label statements. If you want a variable to have a different label, you can change it. If you do not want any labels at all, delete the entire block, resulting in...

```
FILE HANDLE ex /name='C:\HRS\WAVE1\DATA\ex.da' LRECL=15.
```

```
DATA LIST FILE=hhlist /
```

```
  V1 1-5
```

```
  V3 11-15
```

```
  .
```

```
RECODE V1 (99999=SYSMIS).
```

```
RECODE V3 (0=SYSMIS).
```

```
execute.
```

## APPENDIX D: CONCEPTS OF MERGING

### .....D-1 Introduction

Almost any time data needed for a single analysis is found in two separate files, a process called "merging" will need to be done. Merging is a way of combining two different files into one so that their data may be used together.

Merging would not be a particularly difficult task if all datasets were structured the same. Unfortunately, many datasets are not structured alike, and so merging becomes one of the more difficult data management tasks facing the analyst.

A single merging example will follow. The example constructs a "family" file from an individual-level file of respondents and a household-level file of children. The cases given in the example are not actual HRS cases. The variables V1, V2, and V3 (Case ID, Household ID, and Match ID) are the main identification variable types from HRS Wave I. The rest of the variables are similar to actual HRS variables, but were simplified for demonstration purposes.

#### .....D-2 Family Descriptions

In the example that will be used throughout this document, there appears data for four households, which will be described below.

Bob.

Bob is single. He has two daughters, Sarah and Amy.

Jill and Sally.

Jill and Sally are two females in a domestic partnership. Together, they adopted a child named Billy.

Kurt and Paula.

Kurt and Paula are married. While Kurt completed the interview, Paula refused to do so, and so Kurt was asked the questions in Section E. Paula has a son, Ty, which is Kurt's stepson.

Jorge and Maria.

Jorge and Maria are married, but have no kids.

#### .....D-3 Merging Example

##### .....D-3-a Initial files

Our example begins with two files that we wish to

merge: an individual-level file of respondents and a household-level file of their children (if any). The individual-level file has one Respondent per line/record, and the household-level file has one child per line/record. The desired result, upon merging, is a single "family" file with one family per line/record.

#### INDIVIDUAL-LEVEL FILE

| V1    | V2    | V3    | V10   | V11 | V12 |
|-------|-------|-------|-------|-----|-----|
| 620   | 55000 | 0     | Bob   | 1   | 1   |
| 13101 | 12000 | 27102 | Jill  | 1   | 2   |
| 15001 | 46999 | 32999 | Kurt  | 1   | 1   |
| 17100 | 30100 | 20320 | Jorge | 1   | 2   |
| 20320 | 30100 | 17100 | Maria | 2   | 1   |
| 27102 | 12000 | 13101 | Sally | 2   | 1   |

#### HOUSEHOLD-LEVEL FILE

| V2    | V200  | V201 | V202 |
|-------|-------|------|------|
| 12000 | Billy | 2    | 2    |
| 46999 | Ty    | 2    | 1    |
| 55000 | Sarah | 1    | 0    |
| 55000 | Amy   | 1    | 0    |

.....D-3-b Individual-Level File Codebook

VAR #

-----

1 Case ID: Wave I

---

|             |                               |
|-------------|-------------------------------|
| 00001-09999 | Single                        |
| 10000-19999 | Primary Respondent in Pair    |
| 20000-29999 | Secondary Respondent in Pair  |
| 30000-39999 | Refusal by Respondent in Pair |

2 Household ID: Wave I

3 Match ID: Wave I

|                                          |                                                                 |
|------------------------------------------|-----------------------------------------------------------------|
|                                          | 00000 No Secondary Respondent<br>(Primary Respondent is single) |
| 10                                       | Respondent's First Name                                         |
| 11                                       | Respondent Type                                                 |
|                                          | 1 Primary Respondent (R1)<br>2 Secondary Respondent (R2)        |
| 12                                       | Section E Respondent?                                           |
|                                          | 1 Yes<br>2 No                                                   |
| .....D-3-c Household-Level File Codebook |                                                                 |
| VAR #<br>-----                           |                                                                 |
| 2                                        | Household ID: Wave I                                            |
| 200                                      | Child's First Name                                              |
| 201                                      | Relation to Section E Respondent                                |
|                                          | 1 Child<br>2 Adopted Child<br>3 Stepchild                       |
| 202                                      | Relation to Non-Section E Respondent                            |
|                                          | 1 Child<br>2 Adopted Child<br>3 Stepchild                       |

.....D-3-d Intermediate step

The intermediate step in the merging process is to create a set of files from the original files which can be easily merged together into the desired final dataset. There are often multiple ways of merging the same data to produce the same results.

Separating the individual-level file. We first separate the individual-level file into a file of Primary Respondents (R1s) and a file of Secondary Respondents (R2s), on the basis of either their Case ID (V1) or Respondent Type (V11). Because the variables V10 (Respondent Name), V11 (Respondent Type) and V12 (Section E Respondent) appear for both respondents, but are different, we renumber them for the R2 file. Variable V10 becomes V110, V11 becomes V111, and V12 becomes V112 (notice we added "100" to each variable number). We also drop V1 and V3 for the R2 file, because it would only lead to repetition of the same information in the final dataset. (V1 and V3 appear for both respondents, but are the same numbers: Case IDs of the R1 and R2).

It turns out that there are four Primary Respondents in the "R1 file" and two Secondary Respondents in the "R2 file".

| R1 FILE |       |       |       |     |     | R2 FILE |       |      |      |
|---------|-------|-------|-------|-----|-----|---------|-------|------|------|
| V1      | V2    | V3    | V10   | V11 | V12 | V2      | V110  | V111 | V112 |
| 620     | 55000 | 0     | Bob   | 1   | 1   | 30100   | Maria | 1    | 1    |
| 13101   | 12000 | 27102 | Jill  | 1   | 2   | 12000   | Sally | 1    | 2    |
| 15001   | 46999 | 32999 | Kurt  | 1   | 1   |         |       |      |      |
| 17100   | 30100 | 20320 | Jorge | 1   | 2   |         |       |      |      |

Separating the household-level file. In order to merge the children onto the end of the proposed "family" file, we need to separate the household-level file into a number of files equal to the maximum number of kids in the dataset. In this case, the most children in any one household is two. Thus, we will separate the kids into two files, "KID1" and "KID2". Three children will appear in the "KID1" file, and one in the "KID2" file. Thus, one household has two children, and two households have one child. (One household also has no children, and so that household number does not appear in the household-level children file). The variables V200, V201, and V202 in file KID1 will have to be renamed V300, V301, and V302 in the KID2 file, because they overlap.

## KID1 FILE

## KID2 FILE

| V2    | V200  | V201 | V202 | V2    | V300 | V301 | V302 |
|-------|-------|------|------|-------|------|------|------|
| 12000 | Billy | 2    | 2    | 55000 | Amy  | 1    | 0    |
| 46999 | Ty    | 2    | 1    |       |      |      |      |
| 55000 | Sarah | 1    | 0    |       |      |      |      |

.....D-3-e Final step

The final step for the merge is to combine the four files you have created (R1, R2, KID1, and KID2) into one family-level file. All four of the files contain a Household ID, V2, and that will be the basis for the merge. We will first merge the R2 file to the R1 file. In pseudo computer language, it looks like this...

IF (R1's Household ID) = (R2's Household ID), then merge

We now have a file with a Primary Respondent and Secondary Respondent (if present) on the same line/record. If there is no Primary or Secondary Respondent, the appropriate questions should contain missing data codes for those cases.

## R1 FILE + R2 FILE

| V1    | V2    | V3    | V10   | V11 | V12 | V110  | V111 | V112 |
|-------|-------|-------|-------|-----|-----|-------|------|------|
| 620   | 55000 | 0     | Bob   | 1   | 1   | *     | *    | *    |
| 13101 | 12000 | 27102 | Jill  | 1   | 2   | Sally | 2    | 1    |
| 15001 | 46999 | 32999 | Kurt  | 1   | 1   | *     | *    | *    |
| 17001 | 30100 | 20320 | Jorge | 1   | 2   | Maria | 2    | 1    |

\* Asterisks indicate missing data codes, where there was nothing to merge

Next, we merge the KID1 file onto the end of our Primary R + Secondary R file, and then the KID2 file onto the end of that. There is space for each family to have up to the maximum amount of kids encountered (two), but if the family does not have that many kids, missing data codes should appear. An abbreviated representation (the middle is chopped out) of the final "family" file that results is shown below. Asterisks represent missing data codes, where there was nothing to merge.

FAMILY FILE

(R1 FILE + R2 FILE + KID1 FILE + KID2 FILE)

| V1    | V2    | V3    | V10   | V112 | V200  | V201 | V202 | V300 | V301 | V302 |
|-------|-------|-------|-------|------|-------|------|------|------|------|------|
| 620   | 55000 | 0     | Bo... | *    | Sarah | 1    | 0    | Amy  | 1    | 0    |
| 13101 | 12000 | 27102 | Ji... | 1    | Billy | 2    | 2    | *    | *    | *    |
| 15001 | 46999 | 32999 | Ku... | *    | Ty    | 3    | 1    | *    | *    | *    |
| 17001 | 30100 | 20320 | Jo... | 1    | *     | *    | *    | *    | *    | *    |

.....D-3-f Codebook for Resulting "Family" File

VAR #

-----

1 Case ID: Wave I

---

|              |                               |
|--------------|-------------------------------|
| 00001-09999  | Single                        |
| 10000-19999  | Primary Respondent in Pair    |
| 30000-39999  | Refusal by Primary Respondent |
| Missing Data | No Primary Respondent         |

2 Household ID: Wave I

3 Match ID: Wave I

---

|              |                                                           |
|--------------|-----------------------------------------------------------|
| 20000-29999  | Secondary Respondent in Pair                              |
| 30000-39999  | Refusal by Secondary Respondent                           |
| Missing Data | No Secondary Respondent<br>(Primary Respondent is single) |

10 Primary Respondent's First Name

11 Respondent Type (Primary Respondent)

---

|              |                         |
|--------------|-------------------------|
| 1            | Primary Respondent (R1) |
| Missing Data | No Primary Respondent   |

12 Section E Respondent? (Primary Respondent)

|       |                                                  |                            |
|-------|--------------------------------------------------|----------------------------|
|       | 1                                                | Yes                        |
|       | 2                                                | No                         |
|       | Missing Data                                     | No Primary Respondent      |
| 110   | Secondary Respondent's First Name                |                            |
| <hr/> |                                                  |                            |
| 111   | Respondent Type (Secondary Respondent)           |                            |
| <hr/> |                                                  |                            |
|       | 2                                                | Secondary Respondent (R1)  |
|       | Missing Data                                     | No Secondary Respondent    |
| 112   | Section E Respondent? (Secondary Respondent)     |                            |
| <hr/> |                                                  |                            |
|       | 1                                                | Yes                        |
|       | 2                                                | No                         |
|       | Missing Data                                     | No Secondary Respondent    |
| 200   | 1st Child's First Name                           |                            |
| <hr/> |                                                  |                            |
|       | Missing Data                                     | No 1st Child (no children) |
| 201   | 1st Child's Relation to Section E Respondent     |                            |
| <hr/> |                                                  |                            |
|       | 1                                                | Child                      |
|       | 2                                                | Adopted Child              |
|       | 3                                                | Stepchild                  |
|       | Missing Data                                     | No 1st Child (no children) |
| 202   | 1st Child's Relation to Non-Section E Respondent |                            |
| <hr/> |                                                  |                            |
|       | 1                                                | Child                      |
|       | 2                                                | Adopted Child              |
|       | 3                                                | Stepchild                  |
|       | Missing Data                                     | No 1st Child (no children) |
| 300   | 2nd Child's First Name                           |                            |
| <hr/> |                                                  |                            |
|       | Missing Data                                     | No 2nd Child               |

301

## 2nd Child's Relation to Section E Respondent

---

1 Child  
 2 Adopted Child  
 3 Stepchild  
 Missing Data No 2nd Child

302

## 1st Child's Relation to Non-Section E Respondent

---

1 Child  
 2 Adopted Child  
 3 Stepchild  
 Missing Data No 2nd Child

### ...D-4 Summary

In essence, merging is basically the same as "cutting and pasting" in a word processor. You separate ("cut") the files you care to combine into pieces that can later be "pasted" together using a merge program. The result of your efforts is a single file with all the data in one place, on one line.

The example given is one of the more difficult merges you will do (a family file creation). Two individual level files can be combined, as long as their variables do not overlap, simply by matching them on Case ID. No renumbering is needed to combine the individual level files on the HRS data set, and there is no "intermediate step" of creating more files.

The difficulty of creating a family file is that you have to complete the intermediate step mentioned, and match on the basis of more than one identifier (both Case ID and Household ID) throughout the process. If you question whether your family file merge was correct, make sure you accounted for the following types of respondents:

1. Single persons
2. Primary Respondents with partners
3. Primary Respondents without a Secondary Respondent
4. Secondary Respondents without a Primary Respondent

Cases three and four would be due to one respondent completing the interview and the other refusing. The refusal

will have a Case ID in the range 30000-39999.

When merging on the household files, like HHLIST, KIDS, SIBLINGS, and PARENTS, make sure that you leave enough space at the end of the file to include up to the maximum amount of persons present. If there are 40 kids (there never are), you need to leave enough space to string the information for 40 kids onto the end of EVERY household.

Take note that you don't need to merge household files just by the Primary/Secondary respondent distinction. You could merge by gender status as well, to examine pairs by sex.

If you're still not clear on the concept of merging, don't worry! It should be well outlined in manuals related to your statistical package of choice. Be patient, and with time and practice, you'll be able to merge the data into any form you choose. While the actual programming will not be elaborated on here, you can merge all the files in the data set using only two identification variables. Household-level files (HHLIST.DA, KIDS.DA, PARENTS.DA, SIBLINGS.DA) connect to either each other or the individual-level files (the remainder of the files) via the Household ID (V2). Secondly, individual-level files may be connected to each other using the Case ID (V1). A variety of structures can be created, and the final, public-release version of the HRS Wave I data will hopefully be accompanied by working SAS programs that will merge the files into the most common configurations.

#### APPENDIX E: HEALTH AND RETIREMENT STUDY STAFF

Health and Retirement Study  
Institute for Social Research  
P. O. Box 1248  
Ann Arbor, MI 48106-1248

FAX: (313) 747-1186

| Name              | Room | E-Mail Address         |
|-------------------|------|------------------------|
| -----             | ---- | -----                  |
| Barfield, Richard | 3253 | richeb@isr.umich.edu   |
| Brown, Charlie    | 3258 | charlieb@isr.umich.edu |
| Hewett, Heather   | 3250 | heatherh@isr.umich.edu |
| Hill, Dan         | 3256 | dhhill@isr.umich.edu   |
| Hill, Martha      | 3100 | hillm@isr.umich.edu    |
| Howell, David     | 3253 | dahowell@isr.umich.edu |

|                   |      |                        |
|-------------------|------|------------------------|
| Juster, F. Thomas | 3240 | ftjuster@isr.umich.edu |
| Keller, Janet     | 3268 | janetpjk@isr.umich.edu |
| Kempton, Dorothy  | 3252 | dorothyk@isr.umich.edu |
| Liebowitz, Cathy  | 3249 | catlieb@isr.umich.edu  |
| Sullivan, Colleen | 3251 | cmsull@isr.umich.edu   |

## APPENDIX F: GLOSSARY OF TERMS

When working with the HRS data and its documentation, a number of terms and their abbreviations will be encountered. A few of the most common are listed below. Please excuse definitions in which other defined terms are used.

**Age-Eligible Respondent:** A respondent born in the years 1931-1941, inclusive.

**Age-Eligible Household:** A household with at least one age-eligible respondent (each HRS household is supposed to have at least one age-eligible respondent, but there are a few exceptions).

**Case ID:** See "case identifier".

**Case Identifier:** [also "Case ID".] A unique five-digit number assigned to each individual respondent. Case IDs from 1-9999 were given to respondents who are single. Primary respondents (R1s) received Case IDs ranging from 10,000 to 19,999, while Secondary respondents (R2s) received Case IDs ranging from 20,000-29,999. Variable one is always the Case ID, and is present in all individual-level files.

**Family Composition:** Refers to persons related to either the R1 or the R2. In the HRS, there are separate files for children, siblings, and parents, whether they live in the household or not. Other relatives are present in the household listing if they live in the household.

**HH:** See "household".

**H/P:** Husband/partner (of a respondent, normally).

**Household:** [also "HH"] A household sampled by the HRS, including all persons who live there.

**Household Identifier:** Each HRS household (up to two respondents and persons living in the household with them) has a unique

five digit Household ID. The Household ID is always found as variable ten, and is present in all files.

Household-Level Files: Data files in which each record/line contains not information of an individual nature, but concerning the household. Household Listing One of the data files; a list of every person living in the household, one person to a record/line.

HRS: Health and Retirement Study. Health and Retirement Survey when referring to the instrument.

Individual-Level Files: Data files in which there is one respondent per record/line.

Line Number: A family member identifier primarily for internal purposes.

Married Pairs: See "pairs".

Match ID: See "match identifier".

Match Identifier: [also "Match ID"] In individual-level files, the Match ID always appears as variable 22. For Singles, the Match ID always contains zeroes (for it is inappropriate). For Pairs, the Match ID contains the Case ID for the Respondent's partner. You may notice that some Match IDs are in the range 30000-39999, which is not a valid Case ID. Numbers in the 30000 range were assigned to households where one respondent completed the interview, and the partner refused. The partner was assigned a Case ID in the 30000 series, but is not listed as a separate record, only in the Match ID of the respondent who completed the interview.

Pairs: [also "Married Pairs"] Households in which there are both a primary and secondary respondent (which are married or partners).

Person Number: A three-digit identifier for each family member. The number is unique within each household (no other family members will have the same person number, but the numbers are reused in other households).

Primary R: See "primary respondent".

Primary Respondent: [also "R1", "primary R"] The primary respondent was to be either the sole respondent in a single respondent household, or the respondent most knowledgeable of the household financial matters in a household with two respondents who were married/partners. The primary respondent answered the sections regarding housing, assets, and liabilities.

R: See "respondent".

R1: See "primary respondent".

R2: See "secondary respondent".

Relationship Code: A code which represents the relationship of each family or household member to the respondents.

Respondent: [also "R"] A person surveyed by the HRS.

Secondary R: See "secondary respondent".

Secondary Respondent: [also "R2", "secondary R"] Whether a respondent is the primary or secondary respondent is really of no value analytically other than to methodologists. A secondary respondent is defined as the respondent in a married/partner pair who is not the primary respondent. (The primary respondent was the only one to answer housing and asset questions and provided answers for both themselves and their partner.)

Sect E R: See "Section E respondent".

Section E R: See "Section E respondent".

Section E Respondent: [also "Section E R", "Sect E R"] The respondent that was asked to answer Section E (one per household). Section E concerned family composition, the household listing, and other family information.

Singles: Households in which there is only one respondent, who is single (may refer either to the individual or the household).