HEALTH AND RETIREMENT STUDY
2004 Core
Final, Version 1.0
August 2006

Data Description and Usage

Data Description and Usage

1. Overview[1]

The 2004 HRS Core (Final, Version 1.0) data release consists of data obtained as part of the Health and Retirement Study (HRS), a national longitudinal study of the economic, health, marital, and family status, as well as public and private support systems, of older Americans. The National Institute on Aging provided funding (NIH U01 AGO9740), with supplemental support from the Social Security Administration. The Institute for Social Research (ISR) Survey Research Center (SRC) at the University of Michigan conducted the survey.

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

1A. The Sample Interviewed in 2004

The data collection period for the 2004 interview was March 2004 through February 2005. The HRS sample is comprised of five sub-samples.

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

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

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

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

The Early Baby Boomer (EBB) sub-sample consists of people who were born in 1948 through 1953, were household residents of the U.S. when first interviewed in 2004, and who, at that time, did not have a spouse or partner who was born

before 1948, and their spouses or partners at the time of the initial interview or at the time of any subsequent interview. The Early Baby Boomer sub-sample was interviewed in 2004 and every two years thereafter.

Original sample members are those selected as described above and their spouses or partners at the time of the initial interview in 1992 (HRS), 1993 (AHEAD), 1998 (CODA or WB) or 2004 (EBB). For more details about the sample, see our Web site.

1B. 2004 Questionnaire Sections

The content of the 2004 data collection instrument is roughly equivalent to the content of the HRS 2002 instrument. However, in 2004 two new sections were added to the instrument: Section I - Physical Measures and Section LB - Leave-Behind Questionnaires.

2004 Section	Content
PR	Preload
A	Coverscreen
B	Demographics
C	Physical Health
D	Cognition
E	Family Structure
F	Parents, Siblings and Transfers
G	Functional Limitations and Helpers
H	Housing
I	Physical Measures
J	Employment
K	Last Job
L	Job History
M	Disability
N	Health Services and Insurance
P	Expectations
Q	Assets and Income
R	Asset Change
S	Widowhood and Divorce
T	Wills and Life Insurance
U	Asset Reconciliation
V	Modules
W	Event History, Internet Use and Social Security
RC	Repeat Cognition
Y	Time Calculations
IO	Interviewer Observations
LB	Leave-Behind Questionnaires

1C. Levels of Files

In the 2004 data collection instrument, most questions were asked of all respondents. Some questions were asked about the household. For two-respondent households, household level questions were asked of one respondent who was designated as the financial respondent, family respondent, or coverscreen respondent (the first respondent interviewed) on behalf of the entire household.

In addition to the familiar household-level and respondent-level files, the 2004 HRS Core (Final, Version 1.0), contains files at five other levels: household-member-and-child, sibling, helper, transfer-to-child and transfer-from-child.

1C1. Household Level Files

Household-level files contain questions that were asked about the household of a designated coverscreen, financial, or family respondent. A coverscreen respondent answered family questions (section A) on behalf of the entire household; the coverscreen respondent may or may not be the family respondent. A family respondent answered family questions (section E) on behalf of the entire household, and a financial respondent answered household-level financial questions (sections H, Q, R and U) on behalf of the entire household. The household-level files contain one record for each household in which at least one interview was obtained in 2004.

1C2. Respondent Level Files

Respondent-level files contain questions that were asked of all respondents about themselves (or asked of a proxy about the respondent if the respondent was unable to give an interview). The files contain one record for each respondent or proxy who gave an interview in 2004.

1C3. Sibling Level File

The sibling file consists of characteristics of the respondent's siblings. If a respondent had at least one living parent, he/she was asked a variety of questions about his/her siblings (Section F). The sibling file contains one record for each sibling of a respondent. Each respondent reports on their own parents and siblings. Sibling data are also stored in the preload section, H04PR_sb.

1C4. Household Member and Child Level Files

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

1C5. Helper Level File

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

1C6. Transfer-to-Child-Level File

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

1C7. Transfer-from-Child-Level-File

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

2. File Naming Conventions

Files are named beginning with "H04" for HRS 2004, followed by a letter (or two) designating the questionnaire section. A separator, "_" and then one or two letters designating the level, follows the section letter designator. The following letters designate the level of the data files:

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

For example, the file H04A_R includes variables from section A (coverscreen) at the respondent level. Alternatively, file H04H_H includes variables from section H (housing) and the data are at the household level.

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

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

One of each of these file types is provided for each of the 37 data files for the 2004 HRS Core (Final, Version 1.0) data release. For example,

```
H04A_R.DA contains respondent data from section A,
H04A_R.SAS contains corresponding SAS program statements,
H04A_R.SPS contains corresponding SPSS program statements,
H04A_R.DO contains corresponding STATA DO statements,
H04A_R.DCT contains corresponding STATA dictionary statements, and
H04A_R.TXT contains the ASCII codebook.
```

3. Data Files

The 2004 HRS Core (Final, Version 1.0) data are distributed in 37 data files. The files are listed below along with the number of cases (N), number of variables (NV), and the primary identifiers (IDS).

The 2004 HRS Core data are provided in ASCII format, with fixed-length records. Use associated SAS, SPSS or STATA program statements to read the data into the analysis package of your choice.

Household level files, IDS=HHID JSUBHH

```
H04PR_H        N=13645    NV=  46
H04A_H         N=13645    NV=  51
H04E_H         N=13645    NV= 124
H04H_H         N=13645    NV= 194
H04Q_H         N=13645    NV= 675
H04R_H         N=13645    NV= 112
```

```
   H04U_H       N=13645   NV=  50


Respondent level files, IDS=HHID PN

  H04PR_R     N=20129  NV= 175
  H04A_R      N=20129  NV=  27
  H04B_R      N=20129  NV=  92
  H04C_R      N=20129  NV= 204
  H04D_R      N=20129  NV= 155
  H04RC_R     N=20129  NV=  92
  H04F_R      N=20129  NV= 207
  H04G_R      N=20129  NV= 133
  H04I_R      N=20129  NV=  66
  H04IO_R     N=20129  NV= 191
  H04J_R      N=20129  NV=1329
  H04K_R      N=20129  NV= 272
  H04L_R      N=20129  NV= 471
  H04LB_R     N=20129  NV= 181
  H04M1_R     N=20129  NV= 282
  H04M2_R     N=20129  NV= 481
  H04N_R      N=20129  NV= 292
  H04P_R      N=20129  NV=  58
  H04S_R      N=20129  NV= 110
  H04T_R      N=20129  NV=  83
  H04V_R      N=20129  NV= 266
  H04W_R      N=20129  NV=  22
  H04Y_R      N=20129  NV=  70



Household member and child level files IDS=HHID JSUBHH OPN

  H04PR_MC    N=76284  NV=  27
  H04E_MC     N=68796  NV=  28


Sibling level files IDS=HHID JSUBHH OPN

  H04PR_SB    N=20588  NV=  23
  H04F_SB     N=10873  NV=  28


Transfer-to-child-level-file IDS=HHID JSUBHH JTC_NDX

  H04E_TC     N= 6206  NV=  30

Transfer-from-child-level-file IDS=HHID JSUBHH JFC_NDX

  H04E_FC     N=  921  NV=  29

Helper level file  IDS=HHID PN OPN

  H04G_HP     N= 5026  NV=  30
```

4. Identification Variables

Identification variables for HRS 2004 are stored in character format.

4A. Primary Identification Variables

Several variables, HHID, JSUBHH, PN, OPN, JTC_NDX or JFC_NDX are used in various combinations to uniquely identify the seven different level datasets that comprise this data release.

4A1. HHID – Household Identifier

In the initial wave of data collection (in 1992 for the HRS sub-sample, in 1993 for the AHEAD sub-sample, 1998 for the WB and CODA sub-samples, and 2004 for the EBB sub-sample) each sample household was assigned a Household identifier. HHID is stable across waves of data collection and uniquely identifies the original household and any households derived from that household in subsequent waves of data collection. HHID has six-digits.

4A2. JSUBHH – 2004 Sub-household Identifier

In combination with HHID, JSUBHH uniquely identifies a household at the time of the 2004 data collection. Sub-household identifiers can be different at each wave. JSUBHH has one-digit. For more information, see Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments.

4A3. PN – Person Number

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

4A4. OPN – Other Person Number

In the 2004 data collection HHID, JSUBHH and OPN uniquely identify another person in the household member and child files; HHID, PN, and OPN uniquely identify another person in the helper or sibling files. OPN has three-digits. See Datasets Including JSUBHH and OPN as Primary identifiers for additional information.

4A5. JTC_NDX or JFC_NDX – 2004 Transfer Number

In combination with HHID and JSUBHH, JTC_NDX or JFC_NDX uniquely identifies transfer to or from a child or grandchild.   JTC_NDX (index for transfers to children) and JFC_NDX (index for transfers from children) have one digit.

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

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

    1) HHID   HOUSEHOLD IDENTIFIER
    2) PN   PERSON NUMBER

Two identifiers uniquely identify records in the

o   household-level datasets:

        1) HHID   HOUSEHOLD IDENTIFIER
        2) JSUBHH   2004 SUB-HOUSEHOLD IDENTIFIER

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

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

Three identifiers uniquely identify records in the
        o   household-member-and-child-level:

        1) HHID   HOUSEHOLD IDENTIFIER
        2) JSUBHH   2004 SUB-HOUSEHOLD IDENTIFIER
        3) OPN   OTHER PERSON NUMBER

Three identifiers uniquely identify records in the
        o   sibling-level datasets:

        1) HHID   HOUSEHOLD IDENTIFIER
        2) PN   PERSON NUMBER (H04PR_SB only)
        3) JSUBHH   2004 SUB-HOUSEHOLD IDENTIFIER (H04F_SB only)
        4) OPN   OTHER PERSON NUMBER

Three identifiers uniquely identify records in the
        o   transfer-to-child-level, and
        o   transfer-from-child-level datasets:

        1) HHID   HOUSEHOLD IDENTIFIER
        2) JSUBHH   2004 SUB-HOUSEHOLD IDENTIFIER
        3) JTC_NDX/JFC_NDX   2004 TRANSFER NUMBER

4C. Secondary Identification Variables

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

Secondary Identification Variables for
        o   Respondent level,
        o   sibling level and
        o   helper level datasets:

        JSUBHH   2004 SUB-HOUSEHOLD IDENTIFIER
        HSUBHH   2002 SUB-HOUSEHOLD IDENTIFIER
        JPN_SP   2004 SPOUSE/PARTNER PERSON NUMBER
        JCSR   2004 WHETHER COVERSCREEN RESPONDENT
        JFAMR   2004 WHETHER FAMILY RESPONDENT
        JFINR   2004 WHETHER FINANCIAL RESPONDENT

Secondary Identification Variables for
        o   household level,

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

```
HSUBHH    2002   SUB-HOUSEHOLD IDENTIFIER
JPN_CS    2004   COVERSCREEN RESP PERSON NUMBER
JPN_FAM   2004   FAMILY RESP PERSON NUMBER
JPN_FIN   2004   FINANCIAL RESP PERSON NUMBER
JPN_NCS   2004   NON-COVERSCREEN RESP PERSON NUMBER
JPN_NFAM  2004   NON-FAMILY RESP PERSON NUMBER
JPN_NFIN  2004   NON-FINANCIAL RESP PERSON NUMBER
```

4D. Datasets Including JSUBHH and OPN as Primary Identifiers

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

Sometimes records with the same HHID and OPN will be separate reports about the
same person. For example, see the example using 2000 data Couple with Children
and Siblings Divorces, where the couple with children divorces, the record where
HHID is 89012, HSUBHH is 1, and OPN is 101 is the report about the child from
the mother, and the record where HHID is 89012, HSUBHH is 2, and OPN is 101 is
the report about the same child from the father.

In addition, there are instances within the Section F Sibling level file where
the same sibling may have follow up questions from both respondents. This
results from some ambiguity involving sibling relationships to each respondent
in a household. It is not always known, or always clarified, which respondent a
is a blood sibling to each record and which is the in law. If you use HHID and
JSUBHH to merge each record in the H04PR_SB file to each line in the H04F_SB
file, you will get a match to each record. Using the PN of each respondent will
make each line unique.

At other times records with the same HHID and OPN will be reports about
different persons. For example, see the example using 2000 data Couple
Divorces, One Respondent Remarries, Both Split-off Households Have New Members.
– where the couple divorces and both split-off households have new members, the
record where HHID is 90123, HSUBHH is 2, and OPN is 151 is for the ex-wife's
mother, while the record where HHID is 90123, HSUBHH is 1, and OPN is 151 is for
the ex-husband's new stepchild.

5. Distribution Files and Directory Structure

5A. Distribution Files

The files are packaged for download from our Web site in two different ways – as
one large .zip file that contains seven smaller .zip files, one .pdf file, and
one .txt file, or the nine smaller files available individually for separate
download. The combined file is h04core.zip.

The individual .zip files for separate download are:

```
    Data file
        h04da.zip containing data files.

    Program statement files
        h04sas.zip containing SAS data descriptors.
        h04sps.zip containing SPSS data descriptors.
        h04sta.zip containing STATA data descriptors.

    Documentation files
        h04cb.zip containing the codebook.
        h04qn.zip containing the questionnaire.
        h04xref.txt – the HRS 2000 through HRS 2004 cross-reference file
        h04dd.pdf - this document.
```

5B. Directory Structure

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

```
Directory                 Contents

c:\hrs2004                Files downloaded from Web site
c:\hrs2004\codebook       Unzipped files from h04cb.zip
c:\hrs2004\data           Unzipped files from h04da.zip
c:\hrs2004\qnaire         Unzipped files from h04qn.zip
c:\hrs2004\sas            Unzipped files from h04sas.zip
c:\hrs2004\spss           Unzipped files from h04sps.zip
c:\hrs2004\stata          Unzipped files from h04sta.zip
```

Decompress the selected .zip files into the appropriate subdirectories. You will need about 350 MB of free space on your storage device to store the 37 .DA files.

6. Program Statements

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

The statement files are named beginning with the same prefix as the corresponding data file. For example, SAS statements in the file H02A_R.SAS go with the H04A_R.DA data file.

6A. Using the Files with SAS

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

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

If the *.SAS file is located in "c:\hrs2004\sas" and the data file is located in "c:\hrs2004\data", you can run the file as is. A SAS system file (*.SD2 or *.SAS7BDAT) will be saved to directory "c:\hrs2004\sas".

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

6B. Using the Files with SPSS

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

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

If the *.SPS file is located in "c:\hrs2004\spss" and the data file is located in "c:\hrs2004\data", you can run the file as is. An SPSS system file (*.SAV) will be saved to directory "c:\hrs2004\spss".

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

6C. Using the Files with STATA

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

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

If the *.DO and *.DCT files are located in "c:\hrs2004\stata" and the data file is located in "c:\hrs2004\data", you can run the .DO file as is.

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

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

7. Documentation

There are several types of documentation available for use with the 2004 HRS Core (Final, Version 1.0) data release. These include a codebook, the 2004 box-and-arrow questionnaire, and a 2000 to 2004 variable cross-reference listing. In addition, 2004 variables have been added to the Online Concordance.

7A. Codebook

The HRS 2004 Codebook is provided as a series of 37 ASCII text files, as well as a file containing all sections. There is a codebook file corresponding to each data file. Each variable has its own codebook entry. The format of the codebook is, for the most part, consistent with the codebooks of 1995, 1996, 1998, and 2000. The most notable difference is with the naming of variables in the 2002 and 2004 data.

7A1. Variable Names

Variable names begin with a letter designating the wave of data collection (J for 2004), followed by the section letter, and numbers after the section letter. For example, JC001 where J=2004, C=section C (physical health), 001 variable number. Variables from the preload section of the instrument will have either JX or JZ as prefix letters. The X indicates a variable that is updated by data collected in later sections of the questionnaire, whereas the Z indicates preloaded data that were not changed by subsequent answers to questions. For example:

```
JX007_R     RESP FAM/FIN TYPE – UPDATED
JZ077_R     PREV WAVE R FIN/FAM TYPE
```

1A1a. Multiple-response and Looped Variables

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

Simple multiple mention variables take the form: (wave prefix) + (section letter) + (variable number) + (mention number). For example, JN219M1 through JN219M5 are 2004 variables from section N with one through 5 mentions.

Variable names for multiple mentions to questions within a loop take the form: (wave prefix) + (section letter) + (variable number) + (loop iteration) + (letter designating mention number). For example, JJ425a1, is a 2004 variable from section J, variable number 425 in the first iteration of the loop, and the first mention.

Simple loop variables (not a multiple mention) have an underscore (_)in their name and a suffix that designates the loop, e.g., JB066_1.

For variables that have a "W" right after the section designator, the variable names are slightly different. Variable names for multiple mentions to questions within a W-loop take the form: (wave prefix) + (section letter) + ("W") + (variable number) + (letter designating loop iteration) + (mention number). For example, JJW006a1, is a 2004 variable from section J, variable number 006 in the first iteration of the loop, and the first mention. Other non-multiple mention variables within this type of loop are named with the letter designating the loop iteration. For example, JJW001a, is variable number 001 in the first iteration of the loop. In sections where "W" loop variables are present, regular loop variables are designated with a letter to indicate loop iteration. This keeps naming consistent within the section. For example, JJ123a is the first iteration of variable JJ123.

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

1A1b. Masked Variables

To protect the confidentiality of the information that respondents provide, a number of variables have been masked or are simply not included in the Final release public dataset. Some of these variables may be made available to analysts as restricted data. See our Web site for details.

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

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

The names of variables that were masked for confidentiality end with the letter "M"; for example, variable JX026M (1ST ADDRESS STATE – MASKED) and variable JB024M (FATHER USUAL OCCUPATION – MASKED).

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

"Other Specify" and "Open End", or questions that are answered with text, are included in Final release data. Similarly, comments made by respondents in the course of the interview are coded for select questions. The original answer recorded by the interviewer is changed if it is determined that the comment changed the substance of the recorded answer. Therefore, there may be small differences in Final release data compared to Final release data.

7A2. Master Code

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

7B. Other Types of Documentation

In addition to this document and the codebook, three additional types of documentation are available.

7B1. 2000 HRS Final Release Data Description

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

7B2. Box and Arrow Questionnaire

The research community has referred to the type of documentation that describes the questions asked in the interview as a "questionnaire". Since the 2004 HRS data were collected using a CAI program, a traditional hard-copy questionnaire was not produced as part of the data collection phase. However, we have provided a version of the traditional box-and-arrow questionnaire to help document the asking sequence of the questions. The box-and-arrow provides fill designations for the question text where the codebook only displays Blaise code.

7B3. 2000 to 2004 Variable Cross-Reference

As mentioned earlier, the names of the variables in the 2002 and 2004 datasets are markedly different from the names of variables in prior waves. A variable cross-reference table has been assembled to facilitate finding variables of interest in the 2004 data. The cross-reference links the 2004 variable name, level, and label, with the 2000 or 2002 equivalent(s). If a 2004 variable has a 2000 or 2002 equivalent, a reference link appears in the 2004 codebook.

8. Additional Notes

Found here are miscellaneous additional notes regarding HRS 2004 Final Data Release, Version 1. If we become aware of additional issues, they will be posted on our Web site in the Data Alerts section.

8A. Eighteen Respondents Dropped from all Data Files

The following eighteen respondents were dropped from all data files because they do not fit the requirements of the 2004 HRS sample.

HHID	PN
502333	010
500075	010
501383	010
501679	010
502587	010
501851	010
501353	010
501321	010
500765	010
500765	020
501129	010
501129	020
501718	010
501718	020
501887	010
501887	020
501777	010
501777	020

8B. Households with No Coverscreen or Family or Financial Respondents

As noted earlier in this document, the HRS was designed to ask most questions of all respondents, and some questions of just a designated coverscreen, or family, or financial respondent on behalf of the household. However, for some

households we did not obtain an interview from a coverscreen respondent, family or financial respondent.

8B1. There were 2 households that did not have a coverscreen respondent; 132 households had no family respondent; 61 households had no financial respondent. The household records for these households contain null values for the missing information.

8B2. Households missing a coverscreen, family or financial respondent can be identified, respectively, by values of "Blank. No coverscreen/family/financial respondent" in the following variables in the household sections:

```
JPN_CS  - 2004  COVERSCREEN RESP PERSON NUMBER
JPN_FAM - 2004  FAMILY RESP PERSON NUMBER
JPN_FIN - 2004  FINANCIAL RESP PERSON NUMBER
```

8C. Cognition Measures

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

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

8D. Unfolding Bracket Variables and Imputations

Typically, a series of unfolding bracket questions followed a lead-in question asking for an amount. If an actual amount was not given, a series of "unfolding" questions were asked. The manner in which the unfolding questions were programmed (Blaise) is different for the 2002 and 2004 data compared to the CAI (SurveyCraft) software used for 1993 through 2000. This change was transparent to the respondents, since exactly the same questions were asked with the new software as would have been asked with the old software; but it did have an implication for the data that were actually stored and also for the data that are released.

Instead of storing the response to each unfolding question, three summary variables were generated: the minimum and maximum values for the amount, given the answers to the unfolding questions, and if the last answer a respondent gave in an unfolding sequence was either "Don't Know" or "Refused," what that answer was. In 2002, if the Respondent said "more than" to the unfolding question with the highest value, then the maximum value was stored as ten times that value. However, in 2004, if the Respondent said "more than" to the unfolding question with the highest value, then the maximum value was stored as 99999996.

For most analysts, those three variables (and in particular, the minimum and

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

For other sequences -- those in which respondents were randomly assigned to one
of three "entry" points for the first unfolding question -- the analyst will
also need to take into account a fourth variable (located in the preload
sections) that specifies the entry point for each respondent. The following
example shows the preload variable (JZ041) and the unfolding sequence that uses
the random entry point from JZ041.

        Example Random Entry Assignment Variable from Preload:

        Preload Variable from the data file H04pr_h:
....
JZ041   UNFOLD ASSIGN - SELF EMPLOYMENT INCOME
         Section: PR Level: Household Type: Numeric Width: 1 Decimals: 0
         CAI: HH.X041_UnfSEmpInc_V
         2000 Link: G111 2002 Link: HZ041

         ....
         4557   1. RANDOM ASSIGNMENT
         4487   2. RANDOM ASSIGNMENT
         4468   3. RANDOM ASSIGNMENT
          134   Blank. No Data From Prior Wave Interview
....
     Unfolding Series from section Q that uses JZ041 to assign respondents an
     entry point
....
JQ016   R INCOME FROM SELF EMPLOYMENT - MIN
Section: Q   Level: Household   Type: Numeric   Width: 6   Decimals: 0
CAI: SecQ.RIncome.Q016_

  Q016_-Q018_ Unfolding Sequence
  Question text: DID it amount to less than $_____, more than $_____, or what?

        PROCEDURES: UNFM_3up, UNFM_2up1down, UNFM_1up2down
          BREAKPOINTS: 5000, 10000, 25000, 100000
          RANDOM ENTRY POINTS: 5000, 10000, 25000
          RANDOM ENTRY POINT ASSIGNMENT: JZ041

User Note: Entry breakpoint for this unfolding sequence was randomly assigned
in JZ041, located in H04PR_H.

....
          169   0. Value of Breakpoint
            8   5000. Value of Breakpoint
           22   5001. Value of Breakpoint
            8   10000. Value of Breakpoint
           21   10001. Value of Breakpoint
            8   25000. Value of Breakpoint
           35   25001. Value of Breakpoint
            4   100000. Value of Breakpoint
           15   100001. Value of Breakpoint
        13356   Blank. INAP (Inapplicable)

```
================================================================================
JQ017    R INCOME FROM SELF EMPLOYMENT - MAX
Section: Q    Level: Household     Type: Numeric    Width: 8   Decimals: 0
CAI: SecQ.RIncome.Q017_

...............................................................................
            45         4999. Value of Breakpoint
             8         5000. Value of Breakpoint
            21         9999. Value of Breakpoint
             8        10000. Value of Breakpoint
            17        24999. Value of Breakpoint
             8        25000. Value of Breakpoint
            28        99999. Value of Breakpoint
             4       100000. Value of Breakpoint
           151     99999996. Greater than Maximum Breakpoint
         13356         Blank. INAP (Inapplicable)


================================================================================
JQ018    R INCOME FROM SELF EMPLOYMENT - RESULT
Section: Q    Level: Household     Type: Numeric    Width: 2   Decimals: 0
         CAI: SecQ.RIncome.Q018_
...............................................................................
             1           97. Data Not Available
            27           98. DK (Don't Know); NA (Not Ascertained)
           114           99. RF (Refused)
         13504         Blank. INAP (Inapplicable)
...............................................................
```

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

The '97. Data Not Available' code was added to the Result variables in the
unfolding sequences in order to maintain proper sequence flow when a change was
made to the data as a result of coding respondent comments.

The normal operations of the unfolding sequences is that the respondent is asked
a number of questions to determine the maximum and minimum range of a value,
when they respond with DK (Don't Know) or RF (Refuse) at the question where an
amount is asked.

However, sometimes a respondent comment warrants a change of DK or RF to
the amount variable and since the R did not originally go through the unfolding
variables, these data are unknown. Cases such as this will have a '97. Data Not
Available' value in the Result variable of the unfolding sequence.

8E. Modules for the 2004 Data Collection

There are 12 modules for 2004 HRS. Topical areas of the modules include:

Module 1: Risk Aversion
Module 2: Annuities
Module 3: Secondary Causes of Disability
Module 4: Asset Ownership
Module 5: Probability Alternative to Bracketing
Module 6: Loneliness

Module 7: Pension Characteristics and Pension Document Request
Module 8: Retirement Planning
Module 9: Norms on Transfer Behavior
Module 10: Occupational Health
Module 11: Arthritis and Health Behaviors
Module 12: Number Series

8F. Rules - conditions

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

8G. Explanation of the difference between Tracker and 2004 Final data

There are some areas where the tracker file and the 2004 Final release data are different. Below is a list of these situations.

     JSUBHH – The tracker file assigns a '9' for all new spouses in 2004 for the previous wave's SUBHH number. The 2004 Final Release data, however, will list the SUBHH number of the household that the new spouse married into, since the new spouse will 'inherit' many of the relationships from that household.

     JPN_SP – There will be a spouse PN in the ID variables for a respondent, even if that spouse did not give an interview in 2004, and they will be referenced in the household sections by their JPN_SP. Also, for a non-original R, we assign a spouse PN even though we will never interview that person, and track the spouse's influence on the household with that number. The tracker file will not reference these spouses.

     Family and Financial R assignments – As we keep a record of a non-interviewed spouse in the core data, we also release their assigned role in the household (i.e. family or financial R) with a listing for JPN_FIN, JPN_FAM in the household sections. The tracker file will simply list the assignments of the people interviewed. Therefore, for example, the 2004 Core Final data has a blank line in section H for a household where '020' was the non-interviewed financial R. The tracker file would show the same household as not having a financial R.

     Nursing Home – In the process finalizing the 2004 sample weights, we discovered some incorrect classifications in nursing home status (JA028). The 2004 Core data were not changed in order to preserve the original data and flow through the questionnaire. However, the data in the Tracker file reclassified 77 respondents from in a Nursing Home to not in Nursing Home; 4 cases were changed from not in a Nursing Home to in a Nursing Home (see the Tracker file Data Description for a more complete explanation of how suspicious nursing home cases were identified and the criteria used to reclassify respondents).

Marital Status - It is important to note that marital status, as it was assigned in the 2004 data (JB063) may be different from the marital status variable (xMARST) in the Tracker file (see the Tracker file Data Description for a more complete explanation of how marital status was assigned and the criteria used to reclassify respondents). The marital status variable in the Tracker file was constructed by looking at several different sources of data, whereas the marital status variable in the 2004 Core Final data was not changed, in order to preserve the flow through the instrument for respondents.

8H. Changes in Data because of Respondent Comments

The interviewer recorded comments the respondent made during the course of the interview. HRS staff reviewed these comments for selected questions. The coded answer was changed if it was determined that the comment changed the substance of the recorded answer. Occasionally additional codes were added to an existing codeframe. The text of added codeframes appears in the codebook in upper and lower case. Double reporting of income or assets in Section J was corrected wherever possible.

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

8I. Module Assignment Rules

Several rules were used to assign the preload values in JZ009_R (R Module Assignment) and JZ010_R (Spouse Module Assignment). The variables JZ009_R and JZ010_R are in the data file H04PR_R.

> 1. All respondents who were assigned Module 6 last wave will also be assigned Module 6 in 2004.
>
> 2. All re-interview respondents who were not assigned Module 6 were randomly assigned to Modules 1-5, 8-12.
>
> 3. All new respondents (new cohort, new spouse) were randomly assigned to Modules 1-12, and if they were assigned Module 6 or 7 they did not get asked the module this wave.
>
> 4. Two members of a couple were not assigned to the same module and couples were not assigned to receive the combination of modules 2 and 3; 4 and 5; 6 and 7; and 10 and 11.

8J. Description of Number Series (Module 12)

The purpose of this module was to broaden the content domain currently assessed in the HRS to include "fluid intelligence." The specific purpose was to see if we could achieve reasonably informative test scores by using a small subset of items from the Number Series task adapted from the new WJ III using adaptive testing methodology. More detail about the rationale for, development of, and

administration details of this task is available in the Ofstedal, Fisher, & Herzog (2005) HRS Cognition documentation report: (http://hrsonline.isr.umich.edu/docs/userg/dr-006.pdf).

A random sub-sample of HRS respondents was selected and asked to complete the Number Series module at the end of the standard HRS survey. The Number Series test was designed so that each respondent would be asked no more than six test items (i.e., the number of items that we estimated could be completed in the three-minute time frame). The full information in the pattern of responses provided by each respondent was used to create a score on the W-scale (logit metric) based on the scoring method for the original WJ III test. This variable, called JVNScore, represents the respondent's performance on the task where higher scores indicate higher performance on the task. The W scale is a transformation and combination of the Rasch log ability and easy scales (W=9.1024 b + 500). The procedure for developing the transformation was based on the analysis of actual data representing a wide range of ages. In this metric, there is no need for negative numbers or decimal fractions, and the same set of numbers can be used for expressing both item difficulty and a person's ability. This also means that the difference between a person's ability and item difficulty leads to predictive relationships about performance. For example, given an item score with the same W difficulty, the person has a 50% chance of answering correctly. However, given an item whose W-difficulty is 10 points higher, the person has only a 25% change of answering it correctly. Please refer to the cognition documentation report (http://hrsonline.isr.umich.edu/docs/userg/dr-006.pdf) for more information. Other variables include the starting point within the module (JVNSTART) which ranges from 1-4 (where 1= the lowest possible starting point and 4=the highest possible starting point as determined by level of education and performance on the Serial 7s task), the number of questions asked of each respondent (JVNASK), and the number of items answered correctly (JVNCRRCT).

8K. Psychosocial and Disability Leave-Behind Questionnaires Overview and Sample Design (Section LB_R)

In 2004, HRS added a new feature for data collection in the form of self-administered questionnaires that were left with respondents upon the completion of an in-person Core Interview. These are referred to as Leave-Behind Questionnaires (H04LB_R data file) and electronic versions of the questionnaires are included in the 2004 Final release. The purpose of the Leave-Behind Questionnaire is to collect additional information from respondents without adding to the interview length. Two separate questionnaires on different topics were included in HRS 2004: a) Participant Lifestyle Questionnaire, which included questions on participation in general activities, relationships with others, and views on their life in general as well as specific aspects of their life; and b) Participant Questionnaire on Work and Health, which included a series of disability vignettes in which people with various levels of physical and mental health impairment were described and respondents were asked to indicate the extent to which they think the person described is limited in the kind or amount of work they can do. At the beginning of the vignettes, respondents were asked to rate their own degree of limitation in the kind or amount of work they can do.

The Participant Lifestyle Questionnaire had only one version. The Participant Questionnaire on Work and Health contained a total of 24 different versions: 12 for female respondents and 12 for male respondents. Female respondents received a questionnaire in which the individuals described in the vignettes were all female, and males received a questionnaire in which the individuals described in

the vignettes were all male. The questionnaire contained four vignettes on each of three topics: affect, pain, and cardiovascular disease. The different versions varied according to the order in which the topics were presented. In addition, two different introductions to the vignettes were used, one stating "other than the conditions explicitly mentioned, you should imagine the individual in reasonably good health," and the other stating that "other than the conditions explicitly mentioned, you should imagine the individual is like yourself." These introductions were varied across versions. The combination of two introductions and three topics yields 12 distinct versions for each sex.

Sample Design:

Leave-behind questionnaires were administered to a random sample of respondents who received in-person interviews in HRS 2004. Administration of the vignettes began with the April 27 sample release and continued through the end of the 2004 field period. Questionnaire assignments were made by Primary Sampling Unit (PSU) in such a way as to provide roughly equal numbers of respondents for each of the two questionnaires. A random sample of the non-self-representing PSUs (roughly one-half) were assigned to the Participant Lifestyle Questionnaire and the remaining non self-representing PSUs were assigned to the Participant Questionnaire on Work and Health. In self-representing PSUs, the sample was split so that roughly one-half of the segments within each PSU received each type of questionnaire. In households containing two respondents, both respondents received the same type of questionnaire.

The Participant Questionnaire on Work and Health was targeted to respondents under the age of 75. However, some respondents age 75+ received this questionnaire by mistake. These respondents were left in the sample, but they do not represent a random sample of persons in that age group, and should be removed for analyses requiring a random sample.

8L. Sample Design and Description of Physical Performance Measures (Section I_R)

The 2004 physical performance measures sample covers the full age range of HRS age eligible respondents (individuals born prior to 1954). The questionnaire for Physical Performance Measures is available with the 2004 Final release. However, new spouses in 2004, of existing HRS respondents, were not eligible for physical performance measures. The total target sample was approximately 3,900 and 3,339 respondents completed the physical measures. The sample selection criteria were designed to yield approximately 100 individuals for each year of ages -- 51 through 80 -- with declining numbers for ages 81+.

HRS respondents who were deceased, residing in nursing homes, interviewed by proxies, interviewed over the phone, or born later than 1954 were not eligible for inclusion in the physical performance measures. In addition, for practical and economic reasons, respondents were assigned to physical measures (based on birth year and age intervals noted above), if they lived in particular Primary Sampling Units (PSUs). The PSUs included all of the self-representing PSUs, plus about one half of the non self-representing PSUs. The non self-representing PSUs were randomly selected.

In order to achieve the target sample size at each age, the number of selected respondents at each single year of age was inflated to account for age specific rates of mortality, nursing home residence and proxy response as observed in the

HRS 2002. Provided they met the age criteria, special effort was made to include respondents who participated in the physical performance measures conducted as part of the HRS in 1992.

Physical performance measures were collected throughout the 2004 field period, which began February 5, 2004. Interviewers assigned to the PSUs where physical performance measures were taken, received training and were certified prior to the start of the 2004 field period during in person study specific training in Ann Arbor.

Physical Performance Measures –

The physical performance measures include:

Measure	Respondents
Breathing test (peak flow)	Full sample
Grip strength	Full sample
Time walk (8 ft.)	Age 65+
Height and weight	Sub sample of ~800 individuals age 51+

8M. Summary of Data Model (DM) Changes during 2004 Production

New Data Models are periodically released to interviewers during data collection, or field period. New Data Models consist of changes in the Blaise code in the HRS CAPI/CATI survey instrument. Feedback from interviewers and quality control checks of the production data are used to identify what changes are necessary. Every time a new Data Model is released to interviewers, it is given a "version number" and date. In all of the 2004 Final release data files, there is a variable called, JVDATE, which tells what version of the Data Model was used to interview a given household. The following is a brief summary of the programming changes associated with each new Data Model issued during the 2004 field period.

Data Model 2

Section C:
Addition of IWER instruction at JC118, JC119, JC123, and JC124: IWER: THIS QUESTION REFERS TO CIGARETTE SMOKING ONLY. IF R SMOKES CIGARS OR PIPES, ENTER THE TYPE AND AMOUNT SMOKED IN AN F2 COMMENT.

Section J:
JJ412a/d (MIN CONTRIBUTION-EMP STILL CONTRIBUTE) was asked in Version 1 if JJ396a/d (FORM OF EMPLOYER CONTRIBUTION) = 6 (employer doesn't contribute). However, JJ412a/d should have been skipped when JJ396a/d = 6.

Section M:
JM002(HM002 HEALTH PROB) is supposed to be skipped if JZ219=1 (R REPORTED PW HEALTH CONDITION THAT LIMITS WORK), but it was erroneously asked in Version 1.

JM579 (HM579 COMPANY/ORG) did not list, or display, for the interviewers the Respondent's last job in DM 1. When Respondents answered, or gave the identity of their last job at JM579 interviewers had to select 'Other' and enter information about the job in JM580.

Section N:
JN060_1/_3 (EMPLOYER RETIREE HI COVERAGE AFTER 65). The fill for JN060_1/4
(EMPLOYER RETIREE HI COVERAGE AFTER 65) was incorrect in Version 1. Instead of
asking about insurance coverage after age 65 from the Respondent's current
employer, the fill referred to the Respondent's former employer. This was fixed
in Version 2.

Data Model 3

JA036 (MONTH STARTED LIVING WITH NEW SP/P) and JA037 (YEAR STARTED LIVING WITH
NEW SP/P) were erroneously skipped in Versions 1 and 2 for "partnered" couples
in the Early Baby Boom Cohort (EBB). Data Model 3 allows partnered respondents
to get asked JA036 and JA037.

JF001 (MOTHER ALIVE) was supposed to be skipped if the Respondent's mother lived
with the Respondent. It was erroneously asked in Versions 1 and 2 and corrected
in Data Model 3.

Modules:

In Data Models 1 and 2, Respondents who answered 1=Claim $1000 today in JV063
($1000 NOW OR $1100 IN ONE YEAR) were supposed to have been asked JV064 ($1000
NOW OR $1200 IN ONE YEAR) only, and then skipped to the end. Instead they were
also asked V065 ($1000 NOW OR $1050 IN ONE YEAR). This was fixed in Data Model
3, so that after answering JV064 they were skipped to the end.

The result of this error is that 35 Respondents who answered JV063=1 in Data
Models 1 and 2 were erroneously asked JV065.

Data Model 4

Section B:
EBB Respondents who acquired a new spouse, in between the time when their
household was screened for eligibility and their interview, were asked
extraneous questions in Section B regarding how their current/new marriage ended
and how long it lasted (JB068_1/_4 and JB070_1/_4). This was fixed so that these
Respondents were no longer asked the irrelevant questions.

Section G:
There were several flow problems in Section G. JG050 on difficulty taking
medications was being skipped if the Respondent did not receive help using the
telephone (i.e., if JG049=5). Second, JG044 (difficulty shopping) was skipped
if the Respondent did not receive help preparing meals (i.e., if JG043=5).
Third, if two helpers were listed for managing money and the first one was
'97=not on list', that helper was not asked about in the helper loop (JG070-
JG083), or follow-up questions about helpers. Each of these issues was corrected
in DM Version 4.

Section N:
There was a fairly substantial change made to the N103 Branchpoint.
Specifically, cases with no private plans were being asked the sequence of
questions about which plan paid for hospitalization(s) in error. That is, for
N073, N104, N112, and N186 a value of "27" should have been assigned for some

cases and was not. The skip for the follow-up question regarding the name of
the plan that covered the costs worked correctly.

Section P:
Question JP018 (LIKELIHOOD R WILL WORK AFTER AGE 65) was incorrectly being
skipped for Respondents age 62-64. This has been fixed so that all Respondents
under age 65 (who are currently working) are asked JP018. Also, JP017
(LIKELIHOOD R WILL WORK AFTER AGE 62) was only asked of Respondents who were 62
years old, but it should have been asked when Respondent were less than 62 years
of age. This was fixed in Data Model Version 4.

Section S:
Respondents who reported '0' in several amount questions were asked the 'per
week, month...' follow-up questions. Version 4 of the Data Model skips
Respondents over the 'per' questions at JS004_1/_4, JS014_1/_4, JS021, JS031
when the amount given is = 0.

Section V:
Administration of the "Leave-Behind" questionnaires was activated in the release
of the Version 4 DM. The "Leave Behind" sequence starts at JV900 (LEAVE BEHIND
INSTRUCTIONS) and ends with JLB018 (see data file H04LB_R).

Also, response categories 13=REGULAR RETIREMENT PAYMENTS, 14=ACCOUNT, and
15=COMBINATION PLAN were suppressed (no longer an option) for JV315_1/_2 (LISTED
PLAN NAMES - _1 /_2), prior to Data Model 4.

Modules:
Module 5 --
The fills for questions JV212 (CHANCES RECEIVE LESS THAN AMT JV211) and JV213
(CHANCES RECEIVE MORE THAN AMT JV211) in Module 5 were incorrectly defined. The
random number used for other fields were mistakenly being used for JV212 and
JV213 to create fills that displayed the (slightly) wrong question text. This
was fixed in Version 4.

JV222 was dropped and was replaced by JV228 (details follow). In the Final
release data, JV223a, JV224a, JV225a and JV226a were created using the old
versions of the variables. Old Version (JV222) (IF FIN R: Please remind me,
do/ELSE: Do) you (or your husband/wife/partner) have any checking or savings
accounts or money market funds? The codeframes are YES/NO. The New Version
(JV228)as a consequence of Data Model 4: Now we would like you to think about
how much you and your household spent on everything in the past month. Please
think about all bills such as rent, mortgage loan payments, utility and other
bills, as well as all expenses such as food, clothing, transportation,
entertainment and any other expenses you and your household may have. The
codeframe is now 1. CONTINUE for the new intro/preamble type question.

The question text in the follow-up questions to JV228 (formerly JV222) also
required modification. The following changes were made: JV223 (Old Version) --
(JV223a) If you added up all such accounts right now, what are the chances on a
scale of 0 to 100 that they would amount to [less than $5,000/$5,000 or more]?
New Version (JV223) -- On a scale of 0 to 100, where 0 means absolutely no
chance and 100 means absolutely certain, what are the chances that your total
spending last month was [less than $1,000/$1,000 or more]? JV224: Unfolding
wording and amount was changed from $50,000 (JV224a) to $2,000; JV225: Unfolding

wording and amount was changed from $150,000 (JV225a) to $3,000; JV226: Unfolding" wording and amount was changed from $300,000 (JV226a) to $5,000.

Module 9 --
Respondents who were assigned to Module 9 (most of which has to do with children), but who had no living children, were being asked the Module 9 intro question (may we ask you a few more questions...), but then skipped to the COMPLETE screen. This was fixed so that Respondents with no living children, who were selected for Module 9, were asked the relevant portion of Module 9.

Module 10 --
The assignment for Module 10 (which includes questions about occupational health) was not working properly. A group of Respondents who did not classify themselves as working, but were doing some work for pay were being asked the module introduction (will you answer a few more questions). If they said 'yes' they were skipped to the COMPLETE screen. Data Model Version 4 assigned these Rs to Module 11 (question about health [e.g., arthritis] and health behaviors [e.g., smoking]).

Data Model 5

Section A:
Prior to Data Model Version 5, new cohort Respondents were erroneously skipped out of JA034 (Do you consider yourself separated or married?) when they reported that they are not living with their spouse.

Section Q: JQ078: Unfolding Result Value was being saved in the wrong location (over Max Value in JQ078). This has been fixed. Because of this problem, unfolding range information in the JQ078 data prior to DM5 will have incorrect maximum values in that variable.

Data Model 6

Section J
The response categories at JJ713a/d (EMPLOYER CONTRIBUTION LT/GT 5 PCT) were corrected in the English version of the questionnaire. That is, Spanish code value labels were being displayed to interviewers in the English version of the questionnaire, prior to Data Model version 6.

Section N
The response categories for JN148 (NUMBER TIMES SEEN DOCTOR 20X) in the SPANISH version of the questionnaire were corrected, such that 1=LESS THAN 20 TIMES and 5= MORE THAN 20 TIMES (prior to Data Model version 6 they were incorrectly reversed in the Spanish version). The data have been recoded to match the English version; however, for 7 Spanish respondents there is missing data on JN149 where it appears they should have valid data.

Section U
Section U, Asset Reconciliation, was inadvertently dropped from the questionnaire in Data Model Version 5 and was restored in Version 6.

Section W
   The flow in Section W was corrected so that all new cohort Respondents (self
   interviews) received the Internet question (JW303- REGULAR USE OF WEB FOR

EMAIL) and Social Security permission questions (JW306- SOCIAL SECURITY
INTRODUCTION through JW310- SOC SEC PERMIT).

Data Model 7

Section F
The response option at JF080 (Intro to the sibling roster) was corrected so that
5=No was a valid response.

Section U - Asset Reconciliation: Version 7 incorporates corrections to deal
with unanticipated preload values.

Data Model 8

Section N
An Interviewer instruction was added to clarify that JN032_1/_3 (PRIVATE PLAN 1-
3 HELP PAY REGULAR RX) and subsequent questions refer to the private plan, and
not to Medicare.

Data Model 9

Because the field period would continue into 2005, programming changes were made
to accommodate the date of 2005.

8N. Section-Specific Data and Documentation Issues

Section D:

Due to hardware failures in the field, the word list recall data were lost and
unrecoverable for the following respondents:

HHID	PN
010775	020
013678	040
201577	020
500033	010
500100	010
500482	010
500482	020
500747	010
501138	020
501440	010
502212	010
502630	010

Sections E_TC and Section E_FC:

Prior to 2004, in the transfer level files, there was a specific question with
which you could use to create the OPN variable. In 2002, this variable was
HE076 (E_TC):

....
HE076   WHICH CHILD GIVEN LARGEST AMOUNT

Section: E Level: ToChild   Type: Character   Width: 3   Decimals: 0

CAI Reference: BE_TransToKID.E076_ Ref 2000: G2081

To which child did you give the   / (next)) largest amount?

IWER: IF GRANDCHILD: (Which of your children is the parent of that grandchild?)
....
```
      4197                  041-990. Other Person Number
        14                      992. DECEASED CHILD
       412                      993. ALL CHILDREN EQUALLY
       152                      994. ALL GRANDCHILDREN EQUALLY
        65                      995. ALL CHILDREN AND GRANDCHILDREN EQUALLY
       252                      997. OTHER (SPECIFY)
        22                      998. DK (Don't Know)
        19                      999. RF (Refused)
                             Blank. INAP (Inapplicable)
```
....

The problem in 2004 is that both JE076/JE088 are multiple mention questions. Consequently, it is unclear which variable should be the OPN variable, and to/from which child a transfer was given/received. The following variable is from the 2004 Final Release codebook (the same applies when the transfer is FROM a child):
....
JE076M1/M8 WHICH CHILD GIVEN LARGEST AMOUNT
Section: E   Level: ToChild   Type: Character Width: 3   Decimals: 0
CAI: SecE.KIDTransMain.TransToKid[1].E076_[1] 2000 Link: G2081   2002 Link: HE076

To which child did you (or your [husband\wife\partner]) give the ((next)) largest amount?
   INTERVIEWER: IF GRANDCHILD: (Which of your children is the parent of that grandchild?)
   INTERVIEWER: CHOOSE ALL THAT APPLY

....
```
      5323                  041-990. Other Person Number
         5                      992. DECEASED CHILD
       427                      993. ALL CHILDREN EQUALLY
       129                      994. ALL GRANDCHILDREN EQUALLY
        50                      995. ALL CHILDREN AND GRANDCHILDREN EQUALLY
       257                      997. OTHER (SPECIFY)
                               998. DK (Don't Know); NA (Not Ascertained)
                               999. RF (Refused)
        16                   Blank. INAP (Inapplicable)
```

The same issue occurred in H04E_FC - using variables JE088M1 through JE088M6.

Sections M1_R and M2_R:

If the variable was part of the HRS 2002 data collection instrument, then the 2002 name has been provided in the 2004 label.

In the loops asking about "SSDI, SSI, Workers' Compensation, VA, and Other Program," there is a difference in the ask order of month and year questions. In M1, the month is asked first and then the year, and in M2 the year is asked first and then the month.

9. Obtaining the Data

9A. Registration and Downloading the Data

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

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

9B. Conditions of Use

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

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

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

9C. Publications Based on Data

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

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

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

10. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the 2004 HRS Core (Final, Version 1.0) data. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

10A. HRS Internet Site

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

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

10B. Contact Information

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

   Internet: Help Desk at our Web site

   E-mail: hrsquest@isr.umich.edu

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

   FAX: (734) 647-1186

Appendix

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

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

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

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

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

A value of 7 indicates respondents from split household reunited[2].

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

A1.   Married Couple Stays Married.

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

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

```
Time 1
Household records
      HHID=012345 ASUBHH=0
Respondent records
      HHID=012345 PN=010 ASUBHH=0
      HHID=012345 PN=020 ASUBHH=0

Time 2
Household records
      HHID=012345 CSUBHH=0
```

---

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

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

A2. Couple Divorces.

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

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

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

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

A3. One or Both Respondents Die.

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

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

```
Time 1
Household records
     HHID=034567 ASUBHH=0
Respondent records
     HHID=034567 PN=010 ASUBHH=0
     HHID=034567 PN=020 ASUBHH=0

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

A4. Single Respondent Marries.

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

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

```
Time 1
Household records
     HHID=045678 ASUBHH=0
Respondent records
     HHID=045678 PN=010 ASUBHH=0

Time 2
Household records
     HHID=045678 CSUBHH=0
Respondent records
     HHID=045678 PN=010 CSUBHH=0
     HHID=045678 PN=011 CSUBHH=0
```

A5. Couple Divorces, One Respondent Remarries and Divorces.

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

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

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

```
Time 1
Household records
     HHID=056789 ASUBHH=0
Respondent records
     HHID=056789 PN=010 ASUBHH=0
     HHID=056789 PN=020 ASUBHH=0

Time 2
Household records
     HHID=056789 CSUBHH=1
     HHID=056789 CSUBHH=2
Respondent records
     HHID=056789 PN=010 CSUBHH=2
```

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


Time 3
Household records
        HHID=056789 FSUBHH=1
        HHID=056789 FSUBHH=2
        HHID=056789 FSUBHH=5
Respondent records
        HHID=056789 PN=010 FSUBHH=2
        HHID=056789 PN=011 FSUBHH=5
        HHID=056789 PN=020 FSUBHH=1
```

A6.   Couple Divorces and Marries Again.

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

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

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

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


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


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

A7. Married Couple with Children and Siblings.

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

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

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

A8. Couple with Children and Siblings Divorces.

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

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

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

```
Time 1
Household records
     HHID=089012 ASUBHH=0
Respondent records
     HHID=089012 PN=010 ASUBHH=0
     HHID=089012 PN=020 ASUBHH=0
Household member/child records
     HHID=089012 ASUBHH=0 OPN=101 (child)
     HHID=089012 ASUBHH=0 OPN=102 (child)
     HHID=089012 ASUBHH=0 OPN=051 (her brother)
     HHID=089012 ASUBHH=0 OPN=052 (her brother)
     HHID=089012 ASUBHH=0 OPN=061 (his sister)
     HHID=089012 ASUBHH=0 OPN=080 (friend)

Time 2
Household records
     HHID=089012 CSUBHH=1
     HHID=089012 CSUBHH=2
Respondent records
```

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

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

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

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

```
Time 1
Household records
      HHID=090123 ASUBHH=0
Respondent records
      HHID=090123 PN=010 ASUBHH=0
      HHID=090123 PN=020 ASUBHH=0

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