# HEALTH AND RETIREMENT STUDY 2004 Exit Final, Version 1.0 February 2007

Data Description and Usage

# TABLE OF CONTENTS

TABLE OF CONTENTS	III
DATA DESCRIPTION AND USAGE	1
1. OVERVIEW	1
1A. THE SAMPLE INTERVIEWED IN 2004  1B. 2004 QUESTIONNAIRE SECTIONS  1C. LEVELS OF FILES	2 2 2
1C2.       Household Member and Child Level Files	3 3
2. FILE NAMING CONVENTIONS	3
3. DATA FILES	4
4. IDENTIFICATION VARIABLES	4
4A. PRIMARY IDENTIFICATION VARIABLES	5
5A. DISTRIBUTION FILES AND DIRECTORY STRUCTURE	
5B. DIRECTORY STRUCTURE	
6. PROGRAM STATEMENTS	7
6A. USING THE FILES WITH SAS6B. USING THE FILES WITH SPSS6C. USING THE FILES WITH STATA	7
7. DOCUMENTATION	8
7A. CODEBOOK 7A1. Variable Names 7B. OTHER TYPES OF DOCUMENTATION 7B1. 2000 HRS Final Release Data Description 7B2. Box and Arrow Questionnaire	
8. ADDITIONAL NOTES	10
8A. CHANGES IN DATA RESULTING FROM RESPONDENT COMMENTS8B. UNFOLDING BRACKET VARIABLES AND IMPUTATIONS8C. SUMMARY OF DATA MODEL (DM) CHANGES DURING 2004 PRODUCTION	10

9. OBTAINING THE DATA	13
9A. REGISTRATION AND DOWNLOADING THE DATA	13
9B. CONDITIONS OF USE	14
9C. Publications Based on Data	14
10. IF YOU NEED TO KNOW MORE	14
10A. HRS Internet Site	14
10B. Contact Information	14
11. APPENDIX	16
11A. Examples of Sub-Household and Respondent Person Number and Other Person Number	
Assignments	
11A1. Married Couple Stays Married	
11A2. COUPLE DIVORCES	17
11A3. ONE OR BOTH RESPONDENTS DIE	
11A4. SINGLE RESPONDENT MARRIES	
11A5. COUPLE DIVORCES, ONE RESPONDENT REMARRIES AND DIVORCES	18
11A6. COUPLE DIVORCES AND MARRIES AGAIN	19
11B. Cross Wave Merge Using 2004 HRS Exit Data and 2002 Core data	20
11B1. SAS Example of R to R Merge - 2004 Exit to 2002 Core (for only those in the 2004 Exit data file)	20
11B2. Stata Example of R to R Merge - 2004 Exit to 2002 Core (for only those in the 2004 Exit data)	
11B3. SPSS Example of R to R Merge - 2004 Exit to 2002 Core (for only those in the 2004 Exit data file)	
11C. MERGE OF 2004 EXIT INTERVIEW TO DATA OF SURVIVING SPOUSE OR PARTNER IN 2004 CORE	
11C1. SAS Example of 2004 Exit to 2004 Core surviving spouse Merge	
11C2. Stata Example of 2004 Exit to 2004 Core surviving spouse Merge	
11C3. SPSS Example of 2004 Exit to 2004 Core surviving spouse Merge	

# Data Description and Usage

# 1. Overview

The 2004 HRS Exit (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.

The current release of the 2004 Exit (Final, Version 1.0) has 1,227 respondents. An Exit Interview is attempted with a "proxy informant" for panel members who have died. To the greatest extent possible, proxy informants are knowledgeable about the health, family, and financial situation of the deceased (often the proxy is a widow, widower, or some other family member). The content of the 2004 Exit interview is similar to the 2004 core (or living) interview. As such, the 2004 Exit data are amenable to answering researchers' questions concerning medical care, expenditures, how assets are distributed following death, and family decision-making during the latter part of life.

By receiving the data, which have been freely provided, you agree to use them for research and statistical purposes only, and to make no effort to identify the respondents. In addition, you agree to send us 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 Exit 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 of 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 of 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 of 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) cohort consists of people born in 1948 through 1953. The sample selection criterion for the EBB cohort was consistent with the sampling strategy in previous waves of the HRS. Households containing at least one person born between 1948 and 1953 who was either single or whose spouse/partner was born in 1948 or later were designated for the EBB sample. This restriction on spouse's/partner's year of birth was made because individuals married to or partnered with someone born prior to 1948 are already represented in the HRS. Only one individual, plus his/her spouse or partner, was selected per household. Consistent with the original HRS sample design, the 2004 screening for the EBB cohort was designed to produce a 2:1 oversample of Black and Hispanic respondents in the EBB and MBB cohorts.

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). Cases from the HRS, AHEAD, CODA, and WB are present in the HRS 2004 Exit Data. There are no EBB respondents in the 2004 Exit (Final, Version 1.0) because 2004 is the baseline year for the EBB cohort. For more details about the sample, see our Web site.

## 1B. 2004 Questionnaire Sections

The content of the 2004 Exit data collection instrument is equivalent to the content of the HRS 2002 Exit instrument.

2004 Section Content

PR	Preload
A	Coverscreen
В	Demographics
C	Physical Health
D	Cognition
E	Family Structure and Transfers
G	Functional Limitations and Helpers
J	Employment
N	Health Services and Insurance
T	Wills and Life Insurance
W	Event History
IO	Interviewer Observations
Y	Time Calculations

# 1C. Levels of Files

In the 2004 Exit Questionnaire, most questions were asked of all respondents.

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

# 1C1. Respondent Level Files

Respondent-level files contain questions that were asked of all proxy respondents about the deceased respondent. The files contain one record for each exit proxy who gave an interview in 2004 Exit interview.

## 1C2. Household Member and Child Level Files

These files contain characteristics about household members, children, and siblings. The information comes from the PR\_MC file, and contains one record for each household member, child, or sibling.

# 1C3. Helper Level File

This file contains information provided by each exit proxy about the deceased respondent's helpers. A helper may be a person or organization that was reported by the proxy as providing help with ADLs or IADLs in Section G of the 2004 Exit questionnaire. The helper file contains one record for each helper.

## 1C4. Transfer-to-Child-Level File

This file contains information provided by the exit proxy about transfers of money to a child or grandchild. The file contains one record for each transfer to a child or grandchild. Transfers to children are discussed in Section E of the 2004 instrument.

#### 1C5. Transfer-from-Child-Level-File

This file contains information provided by the exit proxy about transfers of money from children or grandchildren. The file contains one record for each transfer from a child or grandchild. Transfers from children are discussed in Section E of the 2004 instrument.

# 2. File Naming Conventions

Files are named beginning with "X04" for HRS 2004 Exit, followed by a letter (or two) designating the questionnaire section. A separator, "\_", and then one or two letters designating the level follow the section letter designator.

```
R for respondent-level

MC for household-member-and-child-level

HP for helper-level

TC for transfer-to-child-level

FC for transfer-from-child-level
```

For example, the file  $X04A_R$  includes variables from section A (coverscreen) at the respondent level. And file  $X04E_TC$  includes variables from section E at the monetary transfer level.

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

```
.da for data files,
.sas for SAS program statements,
.sps for SPSS program statements,
.do for Stata do statements,
.dct for Stata dictionary statements, and
.txt for codebook files.
```

One of each of these file types is provided for each of the 17 data files for the 2004 HRS Exit (Final, Version 1.0) data release. For example,

```
\rm x04A\_R.da contains respondent data from section A, \rm x04A\_R.sas contains corresponding SAS program statements, \rm x04A\_R.sps contains corresponding SPSS program statements, \rm x04A\_R.do contains corresponding Stata do statements,
```

 ${\tt x04A\_R.dct}$  contains corresponding Stata dictionary statements, and  ${\tt x04A\_R.txt}$  contains the ASCII codebook.

# 3. Data Files

The 2004 HRS Exit (Final, Version 1.0) data are distributed in 17 data files. The files are listed below along with the number of cases (N), number of variables (NV), and the primary identifiers (IDS). The records in the data files are sorted in order by these primary identifiers.

The 2004 HRS Exit 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.

Respondent level files, PRIMARY IDS = HHID PN

```
x04PR R N=1,227
                    NV = 79
        N=1.227
x04AR
                    NV = 97
x04B_R N=1,227
                   NV=81
x04C_R
        N=1,227
                   NV=69
x04D_R
         N=1,227
                   NV=61
x04E_R N=1,227
                   NV = 33
x04G_R N=1,227
                   NV=107
x04J_R N=1,227
                   NV=192
       N=1,227
x04N_R
                   NV = 244
\times 04T_R
         N=1,227
                    NV=456
         N=1,227
x04WR
                    NV=11
x04IO R
         N=1,227
                    NV=118
x04YR
         N=1,227
                    NV=18
```

Household member and child level files, PRIMARY IDS = HHID TSUBHH OPN

```
x04PR MC N=6,776 NV=21
```

Transfer-to-child-level-file, PRIMARY IDS = HHID TSUBHH TTC\_NDX

```
x04E_TC N=283 NV=18
```

Transfer-from-child-level-file, PRIMARY IDS = HHID TSUBHH TFC NDX

```
x04E_FC N=153 NV=19
```

Helper-level-file, PRIMARY IDS = HHID PN OPN

```
x04G_{HP} N=2,493 NV=27
```

# 4. Identification Variables

Identification variables for HRS 2004 Final Exit (Version 1.0) are stored in character format.

# 4A. Primary Identification Variables

Several variables, HHID, TSUBHH, PN, OPN, TTC\_NDX, and TFC\_NDX are used in various combinations to uniquely identify records in the five different level datasets that comprise this data release.

## 4A1. HHID - 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. TSUBHH - 2004 Sub-household Identifier

In combination with HHID, TSUBHH uniquely identifies a household at the time of the 2004 data collection. Sub-household identifiers can be different at each wave due to dissolution or reconstitution of a household (e.g. divorce, separation, marriage, or death). TSUBHH has one digit. In the 2004 Exit data, TSUBHH will be either 3 or 4. For more information, see <a href="Examples of Sub-Household">Examples of Sub-Household</a> and Respondent Person Number and Other Person Number Assignments.

#### 4A3. PN - Person Number

In combination with HHID, PN uniquely identifies a respondent. 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, TSUBHH, and OPN uniquely identify another person in the household member and child files; HHID, PN, and OPN uniquely identify another person in the helper files. OPN has three digits.

# 4A5. TTC\_NDX or TFC\_NDX - 2004 Transfer Index Number

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

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

Two identifiers uniquely identify records in the

- o respondent-level datasets:
  - 1) HHID HOUSEHOLD IDENTIFIER
  - 2) PN PERSON NUMBER

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) TSUBHH 2004 SUB-HOUSEHOLD IDENTIFIER
  - 3) OPN OTHER PERSON NUMBER

Three identifiers uniquely identify records in the

- o transfer-to-child-level
- o transfer-from-child-level:
  - 1) HHID HOUSEHOLD IDENTIFIER
  - 2) TSUBHH 2004 SUB-HOUSEHOLD IDENTIFIER

3) TTC\_NDX/TFC\_NDX 2004 TRANSFER INDEX

## 4C. Secondary Identification Variables

Secondary identification variables include HSUBHH, TSUBHH, and TPN\_SP. The secondary identification variables can be used to link the 2004 Exit data with core data from previous waves, or to link a deceased respondent with data from their surviving spouse or partner.

Secondary Identification Variables for

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

```
HSUBHH 2002 SUB-HOUSEHOLD IDENTIFIER
TPN_SP 2004 SPOUSE/PARTNER PERSON NUMBER
```

Secondary Identification Variables for

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

HSUBHH 2002 SUB-HOUSEHOLD IDENTIFIER

# 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 six smaller .zip files and one .pdf file, or the seven smaller files available individually for separate download. The combined file is x04exit.zip.

The individual files for separate download are:

```
Data file x04da.zip containing data files.
```

Program statement files

x04sas.zip containing SAS data descriptors. x04sps.zip containing SPSS data descriptors. x04sta.zip containing Stata data descriptors.

Documentation files

x04cb.zip containing the codebook. x04qn.zip containing the questionnaire. x04dd.pdf - this document.

## 5B. Directory Structure

While a particular setup is not required for using the 2004 HRS Final Exit files, if the following directory structure is used, then no changes to the path name in the data descriptor files are necessary. If you use a different structure, change the directory references in the distribution files.

Directory	Contents
c:\x2004\	Files downloaded from Web site
c:\x2004\codebook	Unzipped files from x04cb.zip
c:\x2004\data	Unzipped files from x04da.zip

c:\x2004\qnaire	Unzipped	files	from	x04qn.zip
c:\x2004\sas	Unzipped	files	from	x04sas.zip
c:\x2004\spss	Unzipped	files	from	x04sps.zip
c:\x2004\stata	Unzipped	files	from	x04sta.zip

Decompress the selected .zip files into the appropriate subdirectories. You will need about 4.0 MB of free space on your storage device to store the 17 .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  $x04A_R$ .sas go with the  $x04A_R$ .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:\x2004\sas" and the data file is located in "c:\x2004\data", you can run the file as is. A SAS system file (\*.sas7bdat) will be saved to directory "c:\x2004\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:\x2004\spss" and the data file is located in "c:\x2004\data", you can run the file as is. An SPSS system file (\*.sav) will be saved to directory "c:\x2004\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: $\x2004$ stata" and the data file is located in "c: $\x2004$ 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 Exit (Final, Version 1.0) data release. These include a codebook, the 2004 box-and-arrow questionnaire, and the data description.

# 7A. Codebook

The HRS 2004 Exit Codebook is provided as a series of 17 ASCII text files, as well as one file containing all sections. There is a codebook file corresponding to each data file. Each variable has its own codebook entry.

#### 7A1. Variable Names

Variable names begin with a letter designating the wave of data collection (T for 2004 Exit), followed by the section letter, and numbers after the section letter. For example, TC030 where T=2004, C=section C (physical health), 030 variable number. Variables from the preload section of the instrument will have either TX or TZ 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:

```
TX007 RESP FAM/FIN TYPE - UPDATED
TZ007 PREV WAVE R FIN/FAM TYPE

TX065_R COUPLENESS STATUS - UPDATED
TZ066_R PREVIOUS WAVE COUPLENESS - INDIVIDUAL
```

# 7Ala. 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, TC021M1 through TC021M6 are 2004 variables from section C with one through six mentions.

Simple loop variables (without a multiple mention) have an underscore (\_) in their name and a suffix that designates the loop, e.g., TN025\_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, TJW044a1, is a 2004 variable from section J, variable number 044 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, TJW001a is variable number 001 in the first iteration of the loop.

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 were retained.

#### 7Alb. 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 2004 Final (Version 1.0) 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.

The names of variables that were masked for confidentiality end in the letter "M"; for example, variable TX046M (1ST ADDRESS STATE - MASKED) and variable TA126M (REGION FACILITY LOCATED - MASKED). In the 2004 Exit release the following variables are masked:

VARIABLE	LABEL
TX026M	1ST ADDRESS STATE - MASKED
TX030M	2ND ADDRESS STATE - MASKED
TX031M	PREV WAVE 2ND RESIDENCE STATE - MASKED
TA126M	R DIED- STATE - MASKED
TA129M	R DEATH CERTIFICATE- STATE - MASKED
TA133M1M	CAUSE OF DEATH-MASKED- 1
TA133M2M	CAUSE OF DEATH-MASKED- 2
TA068M	REGION FACILITY LOCATED - MASKED
TA076M	CURRENT RESIDENCE STATE - MASKED
TA081M	OTHER RESIDENCE STATE - MASKED
TB003M	STATE BORN - MASKED
TB005M	COUNTRY BORN - MASKED
TB017M	R HIGHEST DEGREE - MASKED
TB024M	FATHERS USUAL OCC-MASKED
TB029M1M	TYPE HISPANIC - MASKED
TB031M	R RACE - MASKED
TB047M	ST/COUNTRY LIVED DURING SCHOOL - MASKED
TB052M	DENOMINATION - MASKED
TC209M1M	DIAGNOSIS OF MEMORY PROBLEM-MASKED-1
TC209M2M	DIAGNOSIS OF MEMORY PROBLEM-MASKED-2
TC108M1M	OTHER MEDICAL CONDITIONS-SP-MASKED-1
TC108M2M	OTHER MEDICAL CONDITIONS-SP-MASKED-2
TN314M1M	WHY ADMITTED - FINAL- 1- MASKED
TN314M2M	WHY ADMITTED - FINAL- 2- MASKED
TT158M	WHICH STATE WILL PROBATED - MASKED

# 7B. Other Types of Documentation

In addition to this document and the codebook, two 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 Exit 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.

# 8. Additional Notes

Found here are miscellaneous additional notes regarding HRS 2004 Exit Final Data Release (Version 1.0). If we become aware of additional issues, they will be posted on our Web site in the Data Alerts section.

# 8A. Changes in Data Resulting from Respondent Comments

Interviewers recorded comments made by respondents during the course of the interview. HRS staff reviewed these comments for selected questions. The original 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 sentence case.

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

# 8B. 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. 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. If the last answer a respondent gave in an unfolding sequence was either "Don't Know" or "Refused," that is stored in the result variable.

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 (TZ084) and the unfolding sequence that uses the random entry point from TZ084.

Example Random Entry Assignment Variable from Preload:

```
Preload Variable from the data file X04pr_r:
TZ084 UNFOLD VALUE - TOTAL O/O/P HOSP/NH
       Section: PR Level: Respondent Type: Numeric Width: 1
       Decimals: 0 CAI: RVARS.Z084_UnfNHHosp_V
       389 1. RANDOM ASSIGNMENT
        397
                  2. RANDOM ASSIGNMENT
        441
                  3. RANDOM ASSIGNMENT
______
Unfolding Series from section N that uses TZ084 to assign respondents an entry
point:
TN107
       AMT PAID O-O-P HOSPITAL COSTS - MIN
       Section: N Level: Respondent Type: Numeric Width: 5
       Decimals: 0 CAI: SecN.HospitalStay.N107_
       N107_-N109_ Unfolding Sequence
       Question text: Did it amount to less than $ , more than
       $ , or what?
       PROCEDURES: UNFM_3UplDown, UNFM_2Up2Down, UNFM_1Up3Down
       BREAKPOINTS: $500, $5,000, $10,000, $20,000, $50,000
       RANDOM ENTRY POINTS: $5,000, $10,000, $20,000
       ENTRY POINT ASSIGNMENT 1, 2 OR EMPTY AT Q125/Z084
95 0. Value of Breakpoint
        11
                500. Value of Breakpoint
                501. Value of Breakpoint
               5000. Value of Breakpoint
        10
                5001. Value of Breakpoint
        14
               10000. Value of Breakpoint
         3
               10001. Value of Breakpoint
        31
         2
              20000. Value of Breakpoint
             20001. Value of Breakpoint 50001. Value of Breakpoint
       1012 Blank. INAP (Inapplicable)
```

\_\_\_\_\_\_ TN108 AMT PAID O-O-P HOSPITAL COSTS - MAX Section: N Level: Respondent Type: Numeric Width: 8 Decimals: 0 CAI: SecN. Hospital Stay. N108 16 499. Value of Breakpoint 500. Value of Breakpoint 11 4999. Value of Breakpoint 10 5000. Value of Breakpoint 19 9999. Value of Breakpoint 10000. Value of Breakpoint 19999. Value of Breakpoint 2 20000. Value of Breakpoint 49999. Value of Breakpoint 1 84 99999996. Greater than Maximum Breakpoint 1012 Blank. INAP (Inapplicable) \_\_\_\_\_\_ TN109 AMT PAID O-O-P HOSPITAL COSTS - RESULT Section: N Level: Respondent Type: Numeric Width: 2 Decimals: 0 CAI: SecN. HospitalStay. N109\_ 97. Data Not Available 98. DK (Don't Know); NA (Not Ascertained) 2 99. RF (Refused)

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

1119 Blank. INAP (Inapplicable)

A summary variable called TVDATE appears on all of the data files except interviewer observations. Section IO is a stand alone data collection instrument that is not integrated with the main data collection program. The variable TVDATE indicates which version of the data model was used to collect data from a given proxy respondent. The different data model versions resulted in modest changes to the data correction instrument, and are summarized below:

Data Model Version 2

#### Section C:

Addition of IWER instruction at TC118 and TC119: IWER: THIS QUESTION REFERS TO CIGARETTE SMOKING ONLY. IF R SMOKES CIGARS OR PIPES, ENTER THE TYPE AND AMOUNT SMOKED IN AN F2 COMMENT.

Data Model Version 3

#### Section A:

TA055. The Fin/Fam assignment that was coded out on a completed (or suspended) case was incorrect in Versions 1 and 2. TA055 was always being assigned to Financial and Family Respondents for existing cohort respondents, unless the IWER specifically changed the assignment to something else at that field. However, we determined that the path was not affected by this incorrect assignment; only the assignment in TA055 and the corresponding display in SurveyTrak.

Data Model Version 4

Section G:

TG047 and TG048 were erroneously being asked of exit proxies, therefore was fixed in Version 4.

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 (TG070-TG083). This has been fixed so that both helpers get into the loop.

#### Section T:

Logic was changed before TT124 to capture all exit respondents who owned 2nd homes (regardless of whether the respondents lived in the 2nd home for 2 or more months of the year). This fix also involved a change in the way the associated preload variable was constructed.

Proxy respondents who reported '993=ALL MY CHILDREN' in TT221M1 - TT221M7 (who was involved in the decision regarding treatment during final days of respondent's life) were being skipped past TT227, which asks who was the main decision maker. This has been fixed so that they now get TT227.

Data Model Version 5

No data model changes associated with this version for the  $\operatorname{Exit}$  proxy respondents.

Data Model Version 6

#### Section N:

Corrected code value labels in Spanish version at TN148 (code labels for responses 1 and 5 were reversed in Spanish version).

Data Model Version 7

No data model changes associated with this version for the  $\operatorname{Exit}$  proxy respondents.

Data Model Version 8

Section N:

 $TN032_1/_3$ : IWER instruction added to clarify that  $TN032_1/_3$  and subsequent questions refer to the private plan, not to Medicare.

# 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, and 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 Exit (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

# 11. Appendix

# 11A. Examples of Sub-Household and Respondent Person Number and Other Person Number Assignments

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

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

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

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

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

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

# 11A1. Married Couple Stays Married.

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

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

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

Time 2
Household records
HHID=012345 CSUBHH=0
Respondent records
HHID=012345 PN=010 CSUBHH=0
HHID=012345 PN=020 CSUBHH=0

# 11A2. Couple Divorces.

Time 1

Time 1

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.

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=1

## 11A3. 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.

```
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

Both Respondents Die in same wave:

Household records

HHID=034567 CSUBHH=3 (in exit interview, first respondent in household to
```

die assigned 3)

## 11A4. 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

## 11A5. 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
```

# 11A6. 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
```

# 11B. Cross Wave Merge Using 2004 HRS Exit Data and 2002 Core data

# 11B1. SAS Example of R to R Merge - 2004 Exit to 2002 Core (for only those in the 2004 Exit data file)

```
/* create sas libraries; change paths to match local file locations*/
libname x04 "c:\x2004\sas";
libname h02 "c:\hrs2002\sas" ;
/*Primary identifiers for cross wave, respondent merging are HHID and PN; Sort
data by these variables */
data X04c_r (keep=hhid pn TC028 TC029);
set x04.x04c_r;
run;
proc sort; by hhid pn; run;
data h02b_r (keep=hhid pn HB019 HB031M);
set h02.h02b r;
run;
proc sort; by hhid pn; run;
/*perform the merge, creating the new data set called "cr" - "in" statement on
data set for which you want to keep all records*/
data cr;
merge x04c_r (in=y) h02b_r;
by hhid pn;
if y=1;
run;
```

# 11B2. Stata Example of R to R Merge - 2004 Exit to 2002 Core (for only those in the 2004 Exit data)

```
use HHID PN HB019 HB031M using h02b_r sort HHID PN save tmp1, replace use HHID PN TC028 TC029 using x04c_r sort HHID PN merge HHID PN using tmp1 drop if _m==2 drop _m sort HHID PN save rdata, replace
```

# 11B3. SPSS Example of R to R Merge - 2004 Exit to 2002 Core (for only those in the 2004 Exit data file)

```
GET

FILE='C:\x2004\spss\x04C_R.sav'.

SORT CASES BY

HHID (A) PN (A).

SAVE OUTFILE='C:\x2004\spss\c_r.sav'.

GET

FILE='C:\hrs2002\spss\H02B_R.SAV'.

SORT CASES BY

HHID (A) PN (A).

SAVE OUTFILE='C:\hrs2002\spss\b_r.sav'.
```

```
MATCH FILES /FILE='C:\hrs2002\spss\B_R.SAV'
/FILE='C:\x2004\spss\c_r.sav'
/IN=source01
/BY hhid pn.

VARIABLE LABELS source01
'Case source is C:\x2004\spss\c_r.sav'.
SELECT IF( source01 = 1).
EXECUTE.
```

# 11C. Merge of 2004 Exit Interview to data of surviving spouse or partner in 2004 Core

# 11C1. SAS Example of 2004 Exit to 2004 Core surviving spouse Merge

```
/* create sas libraries; change paths to match local file locations*/
libname x04 "c:\x2004\sas";
libname c04 "c:\hrs2004\sas" ;
data xcr;
set x04.x04c r;
where TPN_SP ne '';
proc sort; by hhid TPN_SP; run;
data ccr;
set c04.h04c_r;
rename PN=TPN SP;
run;
proc sort; by hhid TPN_SP; run;
data excore04;
merge xcr (in=y) ccr;
by hhid tpn_sp;
if y=1;
run;
```

# 11C2. Stata Example of 2004 Exit to 2004 Core surviving spouse Merge

```
use HHID PN JC002 using c:\hrs2004\stata\h04c_r
rename PN TPN_SP
sort HHID TPN_SP
save tmp1

use HHID PN TPN_SP TC028 TC029 using c:\x2004\stata\x04c_r
drop if TPN_SP == ""
sort HHID TPN_SP

merge HHID TPN_SP using tmp1
drop if _m==2
drop _m

sort HHID TPN_SP
save xcdata
```

# 11C3. SPSS Example of 2004 Exit to 2004 Core surviving spouse Merge

```
FILE='C:\hrs2004\spss\H04C_R.sav'.
  RENAME VARIABLES (PN=TPN_SP).
  SORT CASES BY
  HHID (A) TPN_SP (A) .
SAVE OUTFILE='C:\hrs2004\spss\CCR.sav'.
 FILE='C:\x2004\spss\x04C_R.sav'.
FILTER OFF.
USE ALL.
SELECT IF(TPN_SP > "0").
EXECUTE .
SORT CASES BY
 HHID (A) TPN_SP (A) .
SAVE OUTFILE='C:\x2004\spss\xcr.sav'.
MATCH FILES /FILE='C:\hrs2004\spss\CCR.sav'
/FILE='C:\x2004\spss\xcr.sav'
/IN=source01
/BY HHID TPN_SP.
VARIABLE LABELS source01
'Case source is C:\x2004\spss\xcr.sav'.
SELECT IF( source01 = 1).
EXECUTE.
```