

HEALTH AND RETIREMENT STUDY  
2016 Tracker  
Early, Version 3.0  
May, 2019

Data Description and Usage

## Table of Contents

|  |           |
|--|-----------|
| <b>1. INTRODUCTION</b> .....   | <b>4</b>  |
| <b>2. THE STRUCTURE OF TRACKER 2016</b> .....  | <b>6</b>  |
| 2A. VARIABLE LISTING AND DESCRIPTION.....  | 6         |
| 2B. NAMING CONVENTION FOR THE WAVE-SPECIFIC VARIABLES.....   | 11        |
| 2C. VARIABLE FORMATS.....  | 12        |
| <b>3. PERMANENT CHARACTERISTICS</b> .....  | <b>12</b> |
| 3A. DEMOGRAPHIC INFORMATION .....  | 12        |
| 3B. FIRST INTERVIEW.....   | 12        |
| 3C. OVERLAPS .....   | 12        |
| 3D. ENHANCED FACE-TO-FACE ASSIGNMENTS (EFTFASSIGN) .....   | 13        |
| 3E. ADDITIONAL VARIABLES WITH PERMANENT CHARACTERISTICS.....   | 14        |
| <b>4. ANCILLARY STUDIES AND SUPPLEMENTAL VARIABLES</b> .....   | <b>14</b> |
| 4A. 1999 PILOT MAIL SURVEY (MAIL99) .....  | 14        |
| 4B. 2001 MAIL SURVEYS (CAMS01 AND HUMS01) .....  | 14        |
| 4C. 2003 MAIL SURVEYS (CAMS03 AND DIAB03) .....  | 15        |
| 4D. 2005 MAIL SURVEYS (CAMS05 AND PDS05).....  | 16        |
| 4E. 2007 MAIL SURVEYS (CAMS07, DVSA07, DVSB07, AND PDS07) .....  | 17        |
| 4F. 2009 MAIL SURVEYS (CAMS09, HWB09).....   | 19        |
| 4G. 2011 MAIL SURVEYS (CAMS11, HCMS11) .....   | 19        |
| 4H. 2013 MAIL SURVEYS (CAMS13, VA13, HCNS13).....  | 21        |
| 4I. 2015 MAIL SURVEYS (CAMS15, LHMS15) .....   | 22        |
| 4J. 2017 MAIL SURVEYS (CAMS17, LHMS17SPR) .....  | 22        |
| 4K. INTERNET SURVEYS (INTRNT03, INTRNT06, INTRNT07, INTRNT09, INTRNT11 AND<br>INTRNT13) .....  | 23        |
| 4L. NDI (NSCORE, NMONTH, NYEAR, xNDIFLAG: NO LONGER IN TRACKER) .....  | 25        |
| 4M. DEATH DATES AND LAST KNOWN ALIVE DATE (EXDEATHMO, EXDEATHYR, EXDODSOURCE,<br>KNOWNDECEASEDMO, KNOWNDECEASEDYR, KNOWNDECEASEDSOURCE, LASTALIVEMO,<br>LASTALIVEYR, LASTALIVESOURCE)..... | 26        |
| 4N. ADAMS Wave 1 (ADAMS1) .....  | 26        |
| 4O. PHYSICAL MEASURES (PHYMSR04) .....   | 27        |
| 4P. BIOMARKER DATA SETS (GENETICS06, GENETICS08, GENETICS10, GENETICS12).....  | 27        |
| 4Q. LEAVE-BEHIND QUESTIONNAIRES (DISAB04, PSYSOC04, PSYSOC06, PSYSOC08, PSYSOC10,<br>PSYSOC12, PSYSOC14: NO LONGER IN TRACKER) .....   | 27        |
| 4R. HEALTHY COGNITIVE AGING PROJECT (HCAP), 2016.....  | 28        |
| 4S. VENOUS BLOOD STUDY (VBS), 2016.....  | 28        |
| <b>5. WAVE-SPECIFIC VARIABLES</b> .....  | <b>28</b> |
| 5A. INTERVIEW, SAMPLE AND STUDY STATUS.....  | 29        |
| 5A1. <i>xIWTYPE</i> , <i>xIWWAVE</i> , and <i>xINSAMP</i> .....  | 29        |
| 5A2. <i>xRESCODE</i> .....   | 30        |
| 5A3. <i>xSUBHH</i> , and <i>xSUBHHIW</i> .....   | 30        |
| 5B. INTER-RESPONDENT RELATIONSHIP.....   | 31        |
| 5B1. <i>xCOUPLE</i> , <i>xCOUPID</i> , <i>xMARST</i> , and <i>xLIVARR</i> .....  | 31        |
| 5B2. <i>xFAMR</i> and <i>xFINR</i> .....   | 32        |
| 5B3. <i>xNEWSP</i> .....   | 32        |
| 5B4. <i>xPPN</i> .....   | 33        |
| 5B5. <i>xPROXY</i> .....   | 33        |
| 5C. OTHER WAVE-SPECIFIC VARIABLES.....   | 33        |

|   |           |
|---|-----------|
| 5C1. xALIVE.....  | 33        |
| 5C2. xIWLANG, xIWMODE, xIWMONTH, xIWYEAR, and xAGE.....   | 34        |
| 5C3. xNURSHM.....   | 34        |
| 5D. HRS SAMPLE WEIGHTS .....  | 34        |
| 5D1. xWGTHH, xWGTR, xWGTRNH, xWHY0WGT, xWHY0HWT, and xWHY0RWT.....  | 35        |
| 5D2. xWGTHHE, xWGTR, xWHY0HWTE, xWHY0RWTE.....  | 37        |
| 5D3. Sample Weights for 2004 Physical Measure and Leave-Behind<br>Questionnaire (JPMSELWT, JPMWGTR, JWGTR_PS AND JWGTR_DB).....                   | 37        |
| 5D4. Sample Weights for the 2006 Physical Measures, Biomarkers, and<br>Psychosocial Leave Behind Questionnaire (KPMWGTR, KBIOWGTR and KLBWGTR) .. | 38        |
| 5D5. Sample Weights for the 2008-2014 Physical Measures, Biomarkers, and<br>Psychosocial Leave-Behind Questionnaire.....                          | 39        |
| <b>6. IMPORTANT DATA CHANGES IN EARLY TRACKER 2016.....</b>   | <b>40</b> |
| 6A. NDI VARIABLES REMOVED.....  | 40        |
| 6B. NEW VARIABLES .....   | 40        |
| 6C. TWO LINES REMOVED .....   | 40        |
| <b>7. OBTAINING THE DATA.....</b>   | <b>40</b> |
| 7A. REGISTRATION AND DOWNLOADING THE DATA .....   | 40        |
| 7B. CONDITIONS OF USE.....  | 41        |
| 7C. PUBLICATIONS BASED ON DATA .....  | 41        |
| <b>8. IF YOU NEED TO KNOW MORE.....</b>   | <b>41</b> |
| 8A. HRS INTERNET SITE .....   | 41        |
| 8B. CONTACT INFORMATION.....  | 42        |
| <b>APPENDIX.....</b>  | <b>I</b>  |
| <b>9. MERGING THE TRACKER WITH OTHER HRS DATA.....</b>  | <b>I</b>  |
| 9A. HRS 1992 MERGES.....  | I         |
| 9A1. Retaining the Entire, Original HRS Sample.....   | i         |
| 9A2. Keeping Non-Overlaps Only.....   | iii       |
| 9B. HRS 1998 MERGES.....  | IV        |
| 9C. HRS 2010 MERGES FOR 2012 OVERLAPS.....  | V         |

## Data Description and Usage

### 1. Introduction

The HRS Tracker file is created to facilitate the use of HRS data within and across waves. The file contains one record for every person who was ever eligible to be interviewed in any wave. Each record contains basic demographic information, interview status, and if, when and how an interview was conducted in each wave. Also included are cross-sectional weights and information on inter-respondent relationships, which are vital to almost all substantive analyses of the HRS data.

The current version of the Tracker file (Tracker 2016, Early Version 3.0) contains information for panel cases (AHEAD, HRS, CODA, WB, EBB and MBB) along with the Late Baby Boomer (LBB) cohort enrolled for 2016 data collection. Structurally, Tracker 2016 Early Version 3.0 retains most of the information contained in Tracker 2014 Final Version 2.0, and adds information from the 2016 Core Early release and from 2017 ancillary studies. For more information about variables that were included in the 2014 Final Tracker that are no longer included in the 2016 Tracker, please see [Section 6](#).

This document, complementary to the Tracker file codebook, consists of nine sections. Section 2 describes the basic structure of Tracker 2016. Sections 3, 4, 5 and 6 discuss various issues related to the variables in the Tracker. Sections 7 and 8 of the data description outline instructions for obtaining the data and contacting HRS. Section 9 provides some general instructions on merging the Tracker file with other HRS data.

To report any problems regarding the Tracker file or this document, please contact us via email at [hrsquestions@umich.edu](mailto:hrsquestions@umich.edu).

#### 1A. Updates in Tracker 2016, Early, Version 3.0

- Version 3.0, released in May, 2019, adds respondent and household level weights to the tracker file.

Please note, for these early-release weights HRS undertook a new approach to include respondents who report residing in nursing homes. This resulted in a change to the weights, which is described in [Section 5](#), and also resulted in an expansion of the STRATUM values. Until HRS 2014, STRATUM values ranged from 1 to 56; starting in 2016, the value ranges from 1 to 80. (By design, SECU values are always 1 and 2.)

- For Version 3.0, 6 lines from the newly enrolled Late Baby Boomer cohort have been removed from the Tracker file after we determined they were duplicates.

|                      | Removed HHID PN | Retained HHID PN |
|----------------------|-----------------|------------------|
| Duplicate respondent | 541829 010      | 544747 010       |
| Duplicate respondent | 543440 010      | 540807 010       |
| Duplicate respondent | 544121 010      | 541919 010       |
| Duplicate respondent | 544121 020      | 541919 020       |
| Duplicate respondent | 545844 010      | 541521 010       |
| Duplicate respondent | 546027 010      | 544270 010       |

- [Section 5C3](#) of this document, which describes xNURSHM, has also been revised in this version of the data description.

#### 1B. Updates in Tracker 2016, Early, Version 2.0

Version 2.0, released in April, 2019, incorporated the following updates:

- IMMGYEAR: This variable now provides data for 3,034 foreign-born respondents for whom it was inadvertently set to missing in Version 1.0.
- PALIVE: This variable now has a value of 6 (instead of null) for 13,531 respondents who are no longer in the sample and who had values of 5 or 6 on OALIVE.
- PCOUPLE, PSUBHH and PPPN: For 2016 exit cases (PIWTYPE = 11) the variables PPPN and/or PCOUPLE have been corrected for 514 exit cases. PSUBHH has been changed from 3 to 4 for 40 cases.
- PSUBSOCYY: Section [4Q](#) of this document regarding the PSYSOCYY variables has been updated.
- PIWTYPE: This has been changed from 11 to 15 for one respondent. The respondent's EXDEATH date variables have also been set to null. The death date is contained in the new KNOWNDECEASED variables.

- SCHLYRS: This variable has been changed from 0 to 99 for 508 respondents. This total includes 438 members of the Late Baby Boomer cohort who did not provide a core interview this wave, and an additional 70 respondents who were previously missing on SCHLYRS, provided a core interview this wave, but declined to answer question B014 regarding their education.
- SECU and STRATUM have been corrected for cases where the values did not match between spouses. These mismatch cases all met the criteria STUDY = 51, YRENTER = 2010 and MIWTYPE = 5. For SECU, 61 mismatched lines were corrected, and for STRATUM 140 lines were corrected.
- VBSSUBELIG16: The code frame has been corrected for this variable.
- xWHY0WT: The Why0RWt and Why0HWt variables have been corrected for 2012 and/or 2014 for 4 respondents to indicate non-interview rather than deceased. These respondents had been coded as deceased in earlier versions of tracker, but 2016 field work or core interviews have identified them as still living. (The xAlive variables had already been updated for these respondents.)
- xINSAMP, xRESCODE: This variable has been corrected for 2 respondents for M, N and O waves.
- WTCOHORT: The code book text has been updated for this variable.

## **2. The Structure of Tracker 2016**

### ***2A. Variable Listing and Description***

There are 473 Variables in Tracker 2016. These are listed in Table 1 in the order they appear in the Tracker data and the codebook. The first two variables are primary identifiers, household identifier (HHID) and person number (PN), which together uniquely identifies each HRS panel member. Following the two identifiers is a set of permanent characteristics that do not vary across waves (except to be updated or corrected). These variables either provide some basic demographic information for a respondent, or describe their study membership.

It is important to note that starting in the HRS 2006 Core interview the question wording and format for the self-assessed race questions were changed. In short, respondents were able to

identify themselves as part of more than one racial category. Every respondent was asked the new version of the question. Those who indicated multiple races were then asked, "Do you consider yourself primarily (first mention, second mention, etc.)?" For the 2016 Tracker file, the 2016 Core data were only used to update RACE if this value was missing in the Tracker file. In cases where more than one race was mentioned in the Core interview, the follow-up question about the respondent's primary race was used to calculate RACE.

After the primary identifiers and the permanent characteristics, the supplemental variables are related to the HRS Ancillary studies including Off-Year Mail Surveys, Internet surveys, leave-behind questionnaires, ADAMS and other indicators.

The final group of variables include fifteen sets of wave-specific variables, with one set representing each of the HRS survey waves to date. The waves are differentiated by the leading letter in variable names, from A to P (see Table 2). These wave-specific variables document whether a panel member was in a particular sample/study, whether he or she provided an interview in a particular wave, and if an interview was collected, when and how the interview was conducted. When there are two panel members in a sub-household, the wave-specific variables also describe the relationship between them. Moreover, the variables include cross-sectional weights that are often needed for both household level and respondent level analyses. VERSION is also included in this group, indicating the Tracker file version number.

**Table 1. Variables in the 2016 Tracker (Early, Version 3.0): Listing and Description**

**Group A. Primary Identifiers and Permanent Characteristics**

| <b>Name</b> | <b>Type</b> | <b>Length</b> | <b>Label</b>                               |
|-------------|-------------|---------------|--|
| HHID        | Char        | 6             | Household Identifier                       |
| PN          | Char        | 3             | Person Number                              |
| BIRTHMO     | Num         | 2             | Birthdate: Month                           |
| BIRTHYR     | Num         | 4             | Birthdate: Year                            |
| DEGREE      | Num         | 1             | Highest Degree of Education                |
| EFTFASSIGN  | Num         | 1             | Respondents Assignment for EFTF Interviews |
| FIRSTIW     | Num         | 4             | First Interview: Study Year                |
| GENDER      | Num         | 1             | Gender                                     |
| HISPANIC    | Num         | 1             | Hispanicity Type                           |
| IMMGYEAR    | Num         | 4             | Year Immigrated to the U.S.                |
| MOSFLAG     | Num         | 1             | 2010 Minority Oversample Flag              |

**Group A. Primary Identifiers and Permanent Characteristics, con't.**

| <b>Name</b> | <b>Type</b> | <b>Length</b> | <b>Label</b>                           |
|-------------|-------------|---------------|--|
| OVHHID      | Char        | 6             | Overlap Case: Old HHID                 |
| OVPN        | Char        | 3             | Overlap Case: Old PN                   |
| OVRESULT    | Num         | 1             | Overlap Case: Result Code              |
| RACE        | Num         | 1             | Race/Ethnicity                         |
| SCHLYRS     | Num         | 2             | Number of Years in School              |
| SECU        | Num         | 1             | Sampling Error Computation Unit        |
| STRATUM     | Num         | 2             | Stratum ID                             |
| STUDY       | Num         | 2             | Study Membership                       |
| USBORN      | Num         | 1             | Born in the U.S.                       |
| WTCOHORT    | Num         | 3             | Birth Cohort Used for Creating Weights |
| YRENTER     | Num         | 4             | Year Entered the Sample                |

**Group B. Supplemental Variables**

| <b>Name</b> | <b>Type</b> | <b>Length</b> | <b>Label</b>                               |
|-------------|-------------|---------------|--|
| ADAMS1      | Num         | 2             | ADAMS1 Sample Status                       |
| CAMS01      | Num         | 2             | 2001 CAMS Sample Status                    |
| CAMS03      | Num         | 2             | 2003 CAMS Sample Status                    |
| CAMS05      | Num         | 2             | 2005 CAMS Sample Status                    |
| CAMS07      | Num         | 2             | 2007 CAMS Sample Status                    |
| CAMS09      | Num         | 2             | 2009 CAMS Sample Status                    |
| CAMS11      | Num         | 2             | 2011 CAMS Sample Status                    |
| CAMS13      | Num         | 2             | 2013 CAMS Sample Status                    |
| CAMS15      | Num         | 2             | 2015 CAMS Sample Status                    |
| CAMS17      | Num         | 2             | 2017 CAMS Sample Status                    |
| DIAB03      | Num         | 2             | 2003 Diabetes Sample Status                |
| DVSA07      | Num         | 2             | 2007 DVS-A Sample Status                   |
| DVSB07      | Num         | 2             | 2007 DVS-B Sample Status                   |
| EXDEATHMO   | Num         | 2             | Month of Death                             |
| EXDEATHYR   | Num         | 4             | Year of Death                              |
| EXDODSOURCE | Num         | 1             | Source of Exit Date of Death               |
| GENETICS06  | Num         | 1             | Member of 2006 GWAS Sample                 |
| GENETICS08  | Num         | 1             | Member of 2008 GWAS Sample                 |
| GENETICS10  | Num         | 1             | Member of 2010 GWAS Sample                 |
| GENETICS12  | Num         | 1             | Member of the 2012 GWAS sample             |
| HCMS11      | Num         | 2             | 2011 Health Care Mail Survey Sample Status |
| HCNS13      | Num         | 1             | 2013 Health and Nutrition Sample Status    |



**Group B. Supplemental Variables, con't.**

| <b>Name</b>         | <b>Type</b> | <b>Length</b> | <b>Label</b>                                       |
|---------------------|-------------|---------------|--|
| HCAP_SELECT         | Num         | 1             | HCAP Sample Selection                              |
| HCAP16ELIG          | Num         | 1             | HCAP Age-Eligible                                  |
| HCAP16REL           | Num         | 2             | HCAP Release                                       |
| HCAP16RESP          | Num         | 1             | HCAP Response                                      |
| HWB09               | Num         | 2             | 2009 Health and Well-Being Sample Status           |
| HUMS01              | Num         | 2             | 2001 Human Capital Mail Survey Status              |
| INTRNT03            | Num         | 1             | 2003 Internet Survey Sample Status                 |
| INTRNT06            | Num         | 1             | 2006 Internet Survey Sample Status                 |
| INTRNT07            | Num         | 1             | 2007 Internet Survey Sample Status                 |
| INTRNT09            | Num         | 1             | 2009 Internet Survey Sample Status                 |
| INTRNT11            | Num         | 1             | 2011 Internet Survey Sample Status                 |
| INTRNT13            | Num         | 1             | 2013 Internet Survey Sample Status                 |
| KNOWNDECEASEDMO     | Num         | 2             | Known Deceased - Month                             |
| KNOWNDECEASEDYR     | Num         | 4             | Known Deceased - Year                              |
| KNOWNDECEASEDSOURCE | Num         | 1             | Known Deceased - Source                            |
| LASTALIVEMO         | Num         | 2             | Last Alive - Month                                 |
| LASTALIVEYR         | Num         | 4             | Last Alive - Year                                  |
| LASTALIVESOURCE     | Num         | 1             | Last Alive - Source                                |
| LHMS15              | Num         | 2             | 2015 Life History Mail Survey Sample Status        |
| LHMS17SPR           | Num         | 2             | 2017 Spring Life History Mail Survey Sample Status |
| MAIL99              | Num         | 2             | 1999 Mail out Pilot Survey Status                  |
| PDS05               | Num         | 2             | 2005 Prescription Drug Study Sample Status         |
| PDS07               | Num         | 2             | 2007 Prescription Drug Study Sample Status         |
| PHYMSR04            | Num         | 2             | 2004 Physical Measures Sample Status               |
| VA13                | Num         | 2             | 2013 Veterans Mail Survey Sample Status            |
| VBS16SELECT         | Num         | 1             | 2016 VBS Sample Selection                          |
| VBS16ELIG           | Num         | 1             | 2016 VBS Eligible                                  |
| VBS16CONSENT        | Num         | 1             | 2016 VBS Consent                                   |
| VBS16COMPLETE       | Num         | 1             | 2016 VBS Completion                                |
| VBS16VALID          | Num         | 1             | 2016 VBS Valid Assay Results                       |
| VBS16SUBELIG        | Num         | 1             | 2016 VBS Innovative Subsample Eligible             |

**Group C. Wave-Specific Variables**

| <b>Name</b> | <b>Type</b> | <b>Length</b> | <b>Label</b>   |
|-------------|-------------|---------------|--|
| xAGE        | Num         | 3             | Age at Wave X Interview                              |
| xALIVE      | Num         | 1             | Wave X Vital Status                                  |
| xBIOWGTR    | Num         | 5             | Respondent Weight for the Wave X Biomarker Subsample |
| xCOUPID     | Char        | 6             | Wave X Wave Household ID                             |
| xCOUPLE     | Num         | 1             | Wave X Whether Coupled or Partnered                  |
| xFAMR       | Num         | 1             | Wave X Whether Family Respondent                     |
| xFINR       | Num         | 1             | Wave X Whether Financial Respondent                  |
| xINSAMP     | Num         | 1             | Wave X Sample Status                                 |
| xIWLANG     | Num         | 1             | Wave X Interview Language                            |
| xIWMODE     | Num         | 1             | Wave X Interview Mode                                |
| xIWMONTH    | Num         | 2             | Wave X Interview Month                               |
| xIWTYPE     | Num         | 2             | Wave X Interview Type                                |
| xIWWAVE     | Num         | 1             | Wave X Whether Interviewed in the Wave               |
| xIWYEAR     | Num         | 4             | Wave X Interview Year                                |
| xLBWGTR     | Num         | 5             | Respondent Weight for the Wave X Leave Behind Qnaire |
| xLIVARR     | Num         | 1             | Wave X Living Arrangement Status                     |
| xMARST      | Num         | 1             | Wave X Marital Status                                |
| xNEWSP      | Num         | 1             | Wave X New Spouse-Partner Flag                       |
| xNURSHM     | Num         | 1             | Wave X Nursing Home Status                           |
| xPENFLAG    | Num         | 1             | Wave X HRS Pension Data Flag                         |
| xPMSELWT    | Num         | 9             | Sample Weight for Wave X Physical Measures           |
| xPMWGTR     | Num         | 5             | Respondent Weight Wave X Phys Measures Subsample     |
| xPPN        | Char        | 3             | Wave X Spouse-Partner Person Number                  |
| xPROXY      | Num         | 2             | Wave X Proxy Type Status                             |
| xRESCODE    | Num         | 2,4           | Wave X Result Code                                   |
| xSUBHH      | Char        | 1             | Wave X Sub-Household Identifier                      |
| xSUBHHIW    | Num         | 1             | Wave X Whether SUBHH Interviewed                     |
| xWGTHH      | Num         | 8             | Wave X Weight: Household Level                       |
| xWGTHHE     | Num         | 8             | Wave X Weight: Household Level, 2016 onward          |
| xWGTR       | Num         | 8             | Wave X Weight: Respondent Level                      |
| xWGTR E     | Num         | 8             | Wave X Weight: Respondent Level, 2016 onward         |
| xWGTRNH     | Num         | 8             | Wave X Weight: Nursing Home Resident                 |
| xWGTR PS    | Num         | 9             | Resp Weight for Wave X Psychosocial Qnaire           |
| xWGTR DB    | Num         | 9             | Resp Weight for Wave X Disability Qnaire             |
| xWHYOHWT    | Num         | 1             | Wave X Why Zero Household Level Weight               |

**Group C. Wave-Specific Variables, con't.**

| <b>Name</b> | <b>Type</b> | <b>Length</b> | <b>Label</b>   |
|-------------|-------------|---------------|--|
| xWHYOHWTE   | Num         | 1             | Wave X Why Zero Household Level Weight, 2016 onward  |
| xWHYORWT    | Num         | 1             | Wave X Why Zero Respondent Level Weight              |
| xWHYORWTE   | Num         | 1             | Wave X Why Zero Respondent Level Weight, 2016 onward |
| xWHYOWGT    | Num         | 1             | Wave X Why Zero Weight                               |
| VERSION     | Num         | 1             | Tracker File Version Number                          |

**2B. Naming Convention for the Wave-Specific Variables**

The letters A through P, with the exception of I, indicate the specific HRS waves as shown below. All wave-specific variables share the same naming structures across waves. For example, ANEWSP through PNEWSP are new spouse or partner flags for HRS 1992 through HRS 2016 data waves, according to the table below. This feature has two advantages; first, it allows users to identify and link easily all the wave-specific variables, and second, it gives us the ability to name the variables specific to future HRS data waves in a predictable manner (e.g., the variable for new spouse or partner flag for HRS 2018 will be QNEWSP.)

**Table 2. HRS Data Waves and Wave Indicators**

| <b>HRS Data Wave</b> | <b>Wave Indicator</b> |
|----------------------|-----------------------|
| HRS 1992             | A                     |
| AHEAD 1993           | B                     |
| HRS 1994             | C                     |
| AHEAD 1995           | D                     |
| HRS 1996             | E                     |
| HRS 1998             | F                     |
| HRS 2000             | G                     |
| HRS 2002             | H                     |
| HRS 2004             | J                     |
| HRS 2006             | K                     |
| HRS 2008             | L                     |
| HRS 2010             | M                     |
| HRS 2012             | N                     |
| HRS 2014             | O                     |
| HRS 2016             | P                     |

## **2C. Variable Formats**

Following a protocol used in recent HRS data releases, the ID variables are stored in character format. In Tracker 2016, the ID variables are: HHID, PN, OVHHID (Inter-Study or Inter-Respondent Overlap HHID), OVPN (Inter-study or Inter-Respondent Overlap PN), xCOUPID (Wave-Household ID in Wave x), xPPN (Spouse or Partner Person Number in Wave x), and xSUBHH (Sub-Household ID in Wave x).

Non-ID variables (the remainder of the variables in the Tracker file) are stored in numeric format.

## **3. Permanent Characteristics**

### **3A. Demographic Information**

Basic demographic information, such as birth dates, race, Hispanicity, gender, education, and immigration status, may come from a variety of sources. Most often the information was obtained initially through a respondent's answers when he or she was first interviewed. In other cases, the information may come from a spouse or partner or some other knowledgeable person if the sample person was not interviewed. In some cases, the information may have been revised, either by a spouse or partner in an exit interview, or by HRS staff, based on investigations into conflicting information for the respondent. Efforts have been made to resolve all such conflicts in a reasonable way, but it is possible that users may find some differences between the variables in the Tracker file and other HRS data. The Tracker data should be considered definitive.

### **3B. First Interview**

The variable FIRSTIW contains the year a sample member first provided an interview, including core, exit, self or proxy interviews. This variable differs from a sample member's entry cohort or study membership (STUDY) in that a sample member entering a given entry cohort or study might not have been interviewed the first time he or she was eligible for an interview. The variable YRENTER indicates the year that an individual entered the sample regardless of whether an interview was completed.

As of HRS 2016, there are 1090 sample members who have never provided any interviews (self or proxy, core or exit).

### **3C. Overlaps**

Overlaps refer to cases that have multiple IDs and require special handling in constructing longitudinal files and in merging Tracker data to wave-specific files. The variables HHID and PN reflect the current status of the case, while overlap cases also have a former HHID and PN from a previous wave. These

are provided in the variables OVHHD and OVPN.

There are two types of overlap cases in the HRS data. First, there were a number of original HRS 1992 (Wave 1) households eligible to be either an HRS or AHEAD household. An interview was attempted for all of them in HRS 1992. Afterwards, a random subsampling was performed, resulting in 60% of the cases remaining in HRS and the remainder assigned to AHEAD. We refer to the 134 cases transferred to AHEAD as "HRS inter-study overlap" cases. Among those 134 cases assigned to AHEAD, 110 were actually interviewed in AHEAD, and 24 were not.

Second, there are four cases (as of this version of the Tracker) belonging to what we call "household merge overlap" resulting from intermarriage among respondents who entered the study in separate households.

In one case, a married AHEAD sample member (OVHHID=205906 and OVPN=010) became widowed from his original spouse (HHID=205906 and PN=020), and married another AHEAD sample member (HHID=205864 and PN=010) in HRS 1998, gaining a new identity as HHID=205864 and PN=011. This respondent is identified in the Tracker file and the 1998 released data as 205864 011, but for the 1993 and 1995 data waves he appears in the released data under his OVHHID and OVPN identity.

In another case, an unmarried AHEAD sample member (OVHHID=205399 and OVPN=010) married another AHEAD sample member (HHID=208545 and PN=020) in HRS 2000, gaining a new identity as HHID=208545 and PN=010. This respondent is identified in the Tracker file as 208545 010, but appears in the released data under their OVHHID and OVPN identity for the 1993, 1995 and 1998 data waves.

In 2012, two additional household merge overlaps occurred. In both of these cases, a member of the Mid Baby Boomers (MBB) cohort partnered with another member of the same cohort. Specifically, study member OVHHID = 526934, PN = 010 gained a new identity as HHID = 525520, PN = 020; and study member OVHHID = 529766, PN = 010 gained a new identity as HHID = 520845, PN = 020. These two respondents appear in the 2010 released data under their OVHHID and OVPN identities, and in the Tracker file and the 2012 released data under their new identities.

Instructions on how to deal with the various types of overlaps when merging the Tracker file with the HRS Core data are provided in [Section 9](#).

### ***3D. Enhanced Face-to-Face assignments (EFTFASSIGN)***

Starting in 2004, HRS began administering physical measures and biomarker tests for our respondents. The sample is structured so

that half of the respondents receive the enhanced interview each wave. The variable EFTFASIGN holds the respondents' permanent assignment for enhanced face-to-face rotation from 2006 onward. (See also [40. Physical Measures \(PHYMSR04\)](#) and [4P. Biomarker Data Sets \(GENETICS\)](#) below.)

The enhanced face-to-face interview includes a set of physical performance measures, collection of biomarkers, and a Leave-Behind Questionnaire on psychosocial topics. A random one-half of households were pre-selected for the enhanced face-to-face interview in 2006, with the other half of the sample selected for 2008. From that point on every household has repeated the enhanced face-to-face portion every other wave.

The specific physical measures include blood pressure, breathing test (peak flow), grip strength, timed walk (8 ft.), balance tests (semi-tandem, side-by-side, full tandem), height, weight, waist circumference, saliva (for which DNA was extracted and stored) and dry blood spots (analyzed for Hemoglobin Alc, total cholesterol and HDL cholesterol, and C-reactive protein and cystatin C).

### **3E. Additional Variables with Permanent Characteristics**

The remainder of the permanent characteristics variables are SECU and STRATUM, which are used for analysis of sampling error, and WTCOHORT, which provides the entry-wave birth year assignment that can be used to create household level and respondent level weights.

## **4. Ancillary Studies and Supplemental Variables**

Beginning in 1999, questionnaires on various topics have been mailed to subsamples of the HRS every other year. Indicators for these surveys are in the Tracker file, as described below. Data and documentation for these surveys can be found on the HRS website.

### **4A. 1999 Pilot Mail Survey (MAIL99)**

The 1999 Mail Survey was a pilot study and included questions contained in the core survey. The variable **MAIL99** indicates whether an HRS respondent was selected for the 1999 Mail Survey, and if selected, whether the questionnaire was returned.

### **4B. 2001 Mail Surveys (CAMS01 and HUMS01)**

The *2001 Consumption and Activities Mail Survey (CAMS)* asked respondents about their consumption (household expenditures), time spent on activities, and prescription drugs. The variable **CAMS01** indicates whether a respondent was selected to receive a 2001 CAMS questionnaire, and if selected, whether the questionnaire was returned.

The *2001 Human Capital Mail Survey (HUMS)* asked respondents about college expenses (human capital investments) associated with children who had attended college. The variable **HUMS01** indicates whether a respondent was selected to participate in the 2001 Human Capital Mail Survey, and if selected, whether the questionnaire was returned.

#### **4C. 2003 Mail Surveys (CAMS03 and DIAB03)**

The *2003 Consumption and Activities Mail Survey* asked respondents about their consumption (household expenditures), and time spent on activities. The pool for this subsample consisted of respondents who were alive, who had been selected for the 2001 CAMS, and who were not part of a household where their participation in another ongoing HRS mail survey was not complete. From this pool, if the respondent was eligible for CAMS but not eligible for the other 2003 Mail Survey (Diabetes Study), they were selected for the 2003 CAMS. If the respondent was eligible for CAMS and also eligible for the Diabetes Study, and they were part of a coupled household, all eligible respondents were selected. If the respondent was eligible for CAMS and also eligible for the Diabetes Study, and they were an uncoupled household, half of the eligible respondents were selected randomly for CAMS, and half for the Diabetes Study. The variable **CAMS03** indicates whether a respondent was selected to participate in the 2003 CAMS, and if selected, whether the questionnaire was returned.

The *2003 Diabetes Study* asked respondents about aspects of treatment and self-management of diabetes. The HRS 2003 Diabetes Study also collected a clinical biomarker of glucose control: glycosylated hemoglobin, or HbA1c. There were 3,194 interviewed respondents in the 2002 HRS who reported a diagnosis of diabetes (including respondents whose interviews were given by proxies). Of these, 680 were excluded from the 2003 Diabetes Study because of their participation in the CAMS. This exclusion is random, with the exception that proxy cases from 2000 were ineligible for CAMS but eligible for the Diabetes Study, and so they are represented at slightly higher rates prior to weighting adjustments. Of the 2,514 eligible HRS 2002 participants, 133 were subsequently determined to have died prior to the October 2003 start of the Diabetes Study, and so were ineligible for inclusion in the sample. Of the 2,381 remaining eligible cases, 1,901 returned questionnaires, for a response rate of 79.8%. A total of 1,233 valid blood spots were returned for HbA1c assays. That is 64.9% of those who returned questionnaires, and 51.8% of all eligible cases. The variable **DIAB03** in the Tracker data indicates whether a respondent was selected to participate in the 2003 Diabetes Study, and if selected, whether the questionnaire was returned or the respondent had died prior to the start of the study. The variable is further broken down into

respondents who returned a questionnaire with the HbA1c blood kit and those who did not.

#### **4D. 2005 Mail Surveys (CAMS05 and PDS05)**

The *2005 Consumption and Activities Mail Survey* asked respondents about their consumption (household expenditures), and time spent on activities. The pool for the subsample consisted of respondents who were alive, who had been selected for the 2003 CAMS, and who were not part of a household where their participation in another ongoing HRS survey was not complete. From this pool, if the respondent was eligible for the 2005 CAMS (**CAMS05**) but not eligible for the 2005 Prescription Drug Study (**PDS05**), they were selected for the 2005 CAMS. If the respondent was eligible for CAMS and also eligible for the 2005 PDS, and they were part of a coupled household, both halves of the couple were selected for the 2005 CAMS. If the respondent was eligible for 2005 CAMS and also eligible for the 2005 PDS, and they were an uncoupled household, half of the eligible respondents were selected randomly for CAMS 2005, and half for 2005 PDS. In short, the 2005 CAMS sample consists of everyone who had participated in the 2003 CAMS survey, their spouses and/or partners, and a random subsample of HRS respondents who were not in some other mail survey conducted in 2005.

In coupled households, two versions of the 2005 CAMS questionnaire were sent to the household. One version contained questions about household consumption and time spent on various activities; the other asked only about time spent on activities. The respondent who had been designated for the 2003 CAMS was given the full set of questions (consumption and activities), whereas their spouse/partner was given the booklet that asked only about activities. The variable CAMS05 indicates whether a respondent was selected to participate in the 2005 CAMS, and if selected, whether the questionnaire was returned. This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The *2005 Prescription Drug Study*, was the first of a two-wave mail survey designed to track changes in prescription drug utilization as Medicare Part D, the prescription drug benefit, was phased in. The baseline wave, administered in 2005, was intended to capture prescription drug use, coverage, and satisfaction prior to the implementation of Medicare Part D, as well as awareness of the new drug benefit and available subsidies, sources of information on Part D, and expectations of the impact of Part D on prescription drug cost, coverage, and health.



The study sample was comprised of HRS respondents born in 1942 or earlier (65<sup>th</sup> birthday in 2007), or already covered by Medicare or Medicaid at some time between 2002 and 2004. The sample was drawn from respondents of the 2004 HRS Core Survey, including respondents for whom interviews were obtained by proxy. The new drug benefit under Medicare Part D was expected to have its greatest impact on persons without drug coverage, and persons of low income or wealth who may be eligible for subsidies or "extra help" through the Social Security Administration. Because such persons are a minority of the eligible cases, and have shown lower response rates to past self-administered mail questionnaires, the sample included over samples of these groups. There were 5,654 respondents in the 2004 HRS Core Survey who were eligible for participation in the Prescription Drug Survey. Of the eligible participants, 340 died prior to the October 2005 start of the first wave of the Prescription Drug Study and were ineligible for inclusion in the sample. The variable **PDS05** in the Tracker data indicates whether a respondent was selected to participate in the 2005 Prescription Drug Study, and if selected, whether the respondent returned the questionnaire and/or the medication list.

#### **4E. 2007 Mail Surveys (CAMS07, DVSA07, DVSB07, and PDS07)**

The *2007 Consumption and Activities Mail Survey* asked respondents about their consumption (household expenditures), and time spent on activities. The pool for this subsample consisted of respondents who were alive, and who had been selected for the 2005 CAMS. If the respondent was eligible for CAMS and they were part of a coupled household, both halves of the couple were selected for the 2007 CAMS. In short, the 2007 CAMS sample consists of everyone who had participated in the 2005 CAMS survey and their spouses and/or partners. In coupled households, two versions of the questionnaire were sent to the household. One version contained questions about household consumption and time spent on various activities; the other asked only about time spent on activities. The respondent who had been designated for the 2005 CAMS was given the full set of questions (consumption and activities), whereas their spouse/partner was given the booklet that asked only about activities. The variable **CAMS07** indicates whether a respondent was selected to participate in the 2007 CAMS (n=7,744), and if selected, whether the questionnaire was returned (n=5,612). This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The *2007 Disability Vignette Survey (DVS)*, fielded in the fall of 2007, includes a short sequence of questions about the respondents' own health and disability status, followed by a set of anchoring vignettes. The vignettes provide short descriptions

of people in different states of health, which respondents are asked to rate on the same dimensions and scales as they rated their own health. Two versions of the DVS questionnaire (A and B) were administered to assess question ordering and gender effects. The questions in both versions are essentially the same; however, the gender of the hypothetical actor in the vignettes is varied across versions and the questions are presented in a different order.

The DVS Sample consisted of respondents who had completed a self-interview in the HRS 2006 Core and who (prior to the start of the DVS field period) had not died or requested removal from the sample, and who were not in the HRS 2007 Consumption and Activities Mail Survey (CAMS) or the HRS 2007 Prescription Drug Study (PDS). Of the 5,678 questionnaires mailed in the fall of 2007, 4,639 were returned for a simple response rate of 81.7%. Sample members were randomly assigned either Version A or Version B. The variable **DVSA07** indicates whether a respondent was selected to participate in the DVS Survey Version A (n=2,853), and if selected, whether the questionnaire was returned (n=2,329). The variable **DVSB07** indicates whether a respondent was selected to participate in the DVS Survey Version B (n=2,826), and if selected, whether the questionnaire was returned (n=2,310).

The *2007 Prescription Drug Study* (PDS) was the second of a two-wave mail survey designed to track changes in prescription drug utilization as Medicare Part D was phased in. (See PDS05, above, for a description of the baseline wave variable, and HWB09, below, for information about additional follow up.)

The sample for the 2007 PDS consisted of everyone from the original 2005 PDS who responded to PDS 2005, who gave a core HRS interview in 2006, and was not known to be deceased at the beginning of the field period. The 2007 PDS questionnaire was mailed in October 2007 to the eligible sample of 4,990 respondents. After attempting contact with the sample, we determined that 211 persons had died prior to the start of the field period and were ineligible for inclusion in the sample. Of the remaining 4,779 cases, 3,536 returned questionnaires or completed a telephone interview for a response rate of 74%. The field period concluded in September 2008.

The variable **PDS07** in the Tracker file indicates whether a respondent was selected to participate in the 2007 Prescription Drug Study, and if selected, whether the respondent returned the questionnaire and/or the medication list.

#### **4F. 2009 Mail Surveys (CAMS09, HWB09)**

The 2009 Consumption and Activities Mail Survey asked respondents about their consumption (household expenditures), and time spent on activities. In coupled households, two versions of the questionnaire were sent to the household. One version contained questions about household consumption and time spent on various activities; the other asked only about time spent on activities. The sample for the 2009 CAMS mirrored that of 2005 and 2007. In the fall of 2009, a total of 7,231 questionnaires were mailed. Of the 7,231 questionnaires, 4,954 were the "full" version, and 2,277 were the "partial" version. The response rate is 73.7% for the total sample (across both versions of the questionnaires). The response rate for the "full" questionnaire is 72.4% and for the "partial" version the response rate is 76.5%. The variable **CAMS09** indicates whether a respondent was selected to participate in the 2009 CAMS (n=7,231), and if selected, whether the questionnaire was returned (n=5,330). This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The 2009 Health and Well-Being Study is the third wave of a multi-wave mail survey designed to track changes in prescription drug utilization and enrollment in Medicare Part D. (See Prescription Drug Study 2005 and 2007 for more information about the first two waves.)

The sample for HWB 2009 included everyone from the Prescription Drug Study (PDS) 2005/2007 sample plus a 22% random sample of respondents who: a) gave an interview in 2008, b) were not included in the sample for the Consumption and Activities Mail Survey or already included in the PDS sample, and c) born before 1943, and a 64% random sample of respondents who a) gave an interview in 2008, b) were not included in the sample for the Consumption and Activities Mail Survey or already included in the PDS sample, and c) born in 1943 or later. This sample was designed so that 20% of the HRS sample would not be asked to complete a mail survey in the fall of 2009. The final sample size for HWB 2009 was 7,080 respondents, and the response rate was 75%.

#### **4G. 2011 Mail Surveys (CAMS11, HCMS11)**

The 2011 Consumption and Activities Mail Survey asked respondents about their consumption (household expenditures), and time spent on activities. In coupled households, two versions of the questionnaire were sent to the household. One version contained questions about household consumption and time spent on various activities; the other asked only about time spent on activities.

The 2011 CAMS sample is consistent with the procedures outlined in 2005, 2007, and 2009 (above), however new sample was added to the 2011 wave of CAMS consisting of a subsample of the Middle Baby Boomer cohort first interviewed in 2010. To be eligible for CAMS a Middle Baby Boomer respondent had to have given a 2010 Core HRS interview and not be assigned to the HRS 2011 Internet study. There were 2,339 new cohort cases in the 2011 CAMS sample. In September of 2011, 9,078 questionnaires were mailed to the CAMS sample. As was the case since 2005 (see above), there were 2 versions of the questionnaires for the 2011 CAMS. Of the 9,078 questionnaires, 6,162 were the full version, and 2,916 were the partial version. New cohort respondents in coupled/partnered household were randomly assigned to receive either the "full" or "partial" booklet.

The simple response rate is 71.9% for the total sample (across both versions of the questionnaires). The simple response rate for the "full" questionnaire is 70.9% and for the "partial" version the simple response rate is 74.3%. The 2011 CAMS data file contains data from 6,531 respondents of which 1,613 are new cohort cases. The variable **CAMS11** indicates whether a respondent was selected to participate in the 2011 CAMS, and if selected, whether the questionnaire was returned. This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The *2011 Health Care Mail Survey* (HCMS) contains questions about health care, including access, utilization, policy, and veteran's health services. The National Institute on Aging (NIA) provided funding (U01 AG009740) for the 2011 HCMS. In December 2011, questionnaires were mailed to a subsample of HRS respondents (n=10,234). The field period for the 2011 HCMS was mid December 2011 through the end of May 2012.

The subsample for the HCMS consisted of respondents who had given a 2010 core non-proxy interview on or before September 30, 2011 and who were not in the 2011 CAMS. About 10% of randomly selected households were removed from the sample, and an additional 20% of households were selected from those who gave a core interview and who were not in the 2011 CAMS. In the 20% of households and among coupled households, one respondent was randomly chosen to receive a HCMS questionnaire and the other was excluded from the HCMS sample. The intent of the latter selection criteria was to allow for assessing if inclusion of both spouses or inclusion of one spouse affects participation in the subsequent Core wave of the HRS. The data file for the HCMS contains data for 7,652 respondents. The simple response rate for the HCMS is 75 percent. The variable **HCMS11** indicates

whether a respondent was selected to participate in the survey, and whether a questionnaire was returned.

#### **4H. 2013 Mail Surveys (CAMS13, VA13, HCNS13)**

The *2013 Consumption and Activities Mail Survey* sample is consistent with the procedures outlined in previous CAMS waves. In September of 2013, 8,595 questionnaires were mailed to the 2013 CAMS sample. As was the case since 2005 (see above), there were 2 versions of the questionnaires for the 2013 CAMS. Of the 8,595 questionnaires, 5,914 were the "full" version, and 2,681 were the "partial" version.

The simple response rate is 70% for the total sample (across both versions of the questionnaires). The simple response rate for the "full" questionnaire is 69% and for the "partial" questionnaire the simple response rate is 72%. The variable **CAMS13** indicates whether a respondent was selected to participate in the 2013 CAMS, and if selected, whether the questionnaire was returned. This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The *2013 Veterans Mail Survey* covered topics such as general health, use of Veterans Administration services, experience in the military, and friends from the military. In June 2013, questionnaires were mailed to 3811 HRS respondents who had reported military service in the past. The field period concluded in September 2013. The data file for the 2013 Veterans Mail Survey contains data for 1,871 respondents, for a simple response rate of 49%. The variable **VA13** in the Tracker file indicates whether a respondent was selected to participate and whether a questionnaire was returned.

The *2013 Health Care and Nutrition Study* contained questions about health care access, food purchases, food consumption and nutrition (including vitamins and other supplements). In November 2013, questionnaires were mailed to a subsample of HRS respondents (n=12,418) which consisted of all living HRS respondents who were not included in the 2013 Consumption and Activities Mail Survey (CAMS13). The field period for the 2013 HCNS was late November 2013 through early May 2014.

The data file for the HCNS contains data for 8073 respondents. The simple response rate for the HCNS is 65 percent. The variable **HCNS13** in the Tracker file indicates whether a respondent was selected to participate and whether a questionnaire was returned.

#### **4I. 2015 Mail Surveys (CAMS15, LHMS15)**

The 2015 Consumption and Activities Mail Survey sample is consistent with the procedures outlined in previous CAMS waves. In the fall of 2015, 8,039 questionnaires were mailed to the CAMS sample. As was the case since 2005 (see above), there were 2 versions of the questionnaires for the 2015 CAMS. Of the 8,039 questionnaires, 5,628 were the full version, and 2,411 were the partial version.

The simple response rate is 68% for the total sample (across both versions of the questionnaires). The simple response rate for the full questionnaire is 66% and for the partial version the simple response rate is 70%. The variable **CAMS15** indicates whether a respondent was selected to participate in the 2015 CAMS, and if selected, whether the questionnaire was returned. This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The 2015 Life History Mail Survey contained questions about residential and education history and other activities from birth to age 50. In December 2015, questionnaires were mailed to 11,256 HRS respondents who were not included in the 2015 CAMS sample and who completed their most recent HRS core interview in English. The simple response rate for the 2015 LHMS Mail Survey is 58%. The variable **LHMS15** in the Tracker file indicates whether a respondent was selected to participate and whether a questionnaire was returned.

#### **4J. 2017 Mail Surveys (CAMS17, LHMS17SPR)**

The 2017 Consumption and Activities Mail Survey sample is consistent with the procedures outlined in previous CAMS waves. In the fall of 2017, 7,579 questionnaires were mailed to the CAMS sample. As was the case since 2005 (see above), there were 2 versions of the questionnaires for the 2017 CAMS. Of the 7,579 questionnaires, 5,376 were the full version, and 2,203 were the partial version.

The simple response rate is 67% for the total sample (across both versions of the questionnaires). The simple response rate for the full questionnaire is 65% and for the partial version the simple response rate is 69%. The variable **CAMS17** indicates whether a respondent was selected to participate in the 2017 CAMS, and if selected, whether the questionnaire was returned. This variable also breaks down the difference between respondent and spouse eligibility, and whether or not the questionnaire was returned in both of those groups.

The 2017 Spring Life History Mail Survey contained questions about residential, education and employment history and other

activities from birth to age 50. The 2017 Spring LHMS sample consisted of HRS respondents who were also included in the 2015 CAMS sample, who were still alive in 2017, and whose households had been finalized (as either respondents or refusals) on HRS core interviews by early March 2017. A small number of respondents who were in the sample for the 2016 Healthy Cognitive Aging Project (HCAP) and who had not completed their assessment by early March of 2017 were excluded. New spouses of eligible panel respondents were also included.

In June 2017 questionnaires were mailed to the subsample of HRS respondents described above (n = 5,174). The simple response rate was 74%. The variable **LHMS17SPR** in the Tracker file indicates whether a respondent was selected to participate and whether a questionnaire was returned.

#### ***4K. Internet Surveys (INTRNT03, INTRNT06, INTRNT07, INTRNT09, INTRNT11 and INTRNT13)***

##### ***2003 Internet Survey***

In the fall of 2002, contact letters were sent to 2,696 HRS respondents inviting them to participate in the 2003 Internet Survey. The Internet subsample was drawn from the HRS 2002 core sample, wherein respondents were asked if they had internet access and if they would be willing to participate in an Internet survey. The 2003 Internet Survey questionnaire contained several topical areas including Internet/computers, health problem, disability and work limitations, numeracy items, psychosocial items, expectations, and questions about housing, checking accounts, and stocks.

From the 4,784 respondents to the HRS 2002 who indicated they had internet access, 3,496 expressed a willingness to participate in an internet survey. Of the 3,496 respondents who had internet access and said they were willing to do an internet interview, 2,696 were contacted for the 2003 Internet Survey (244 of them also completed Module 5 of the 2005 HRS core survey about computers and internet use), and 800 of 3,496 respondents were assigned to a control group. A total of 2,696 respondents completed the 2003 Internet Survey for a simple response rate of 81.5%. The simple response rate does not adjust for mortality or respondents lost to follow-up. The variable **INTRNT03** in the Tracker file indicates the status for the 2003 Internet Survey.

##### ***2006 Internet Survey***

Contact letters were sent to 1,920 HRS respondents, inviting them to participate in the 2006 Internet Survey, in March 2006. The 2006 Internet Survey questionnaire contained several topical areas including internet/computers, health and emotional problems, Social Security, numeracy items, psychosocial items,

household composition, expectations, questions about housing/checking accounts, and stocks and prescription drug usage and insurance.

The 2006 Internet sample consists of respondents who participated in the HRS 2003 Internet survey (n=785), non-response respondents from the 2003 Internet survey (n=170) and 965 respondents drawn from those who said they had internet access in the HRS 2004 survey and who were not already selected to participate in an HRS 2005 mail survey. A total of 1,352 respondents completed the 2006 Internet Survey, for a simple response rate of 70.4%. The variable **INTRNT06** in the Tracker file indicates the status for the 2006 Internet Survey.

### ***2007 Internet Survey***

In June 2007, contact letters were sent to 3,721 HRS respondents, inviting them to participate in the 2007 Internet Survey. The 2007 Internet Survey questionnaire contained topical areas including internet/computers, health and emotional problems, Social Security, numeracy items, psychosocial items, household composition, housing, checking accounts, stocks, and prescription drug usage and insurance.

The 2007 Internet sample was drawn from those who reported internet access in the HRS 2006 core survey and whose households were not already selected to participate in another HRS sub sample (ADAMS). From this pool, we selected the 1,369 respondents who were chosen for the HRS 2003 Internet Survey but not the HRS 2006 Internet Survey, and 80% of the 2,940 respondents (n=2,352) who were not chosen for either the HRS 2003 Internet Survey or the HRS 2006 Internet Survey, for a total sample of 3,721. A total of 2,665 respondents completed the 2007 Internet Survey, for a simple response rate of 71.6%. The field period for this survey ran from June 2007 through October 2007. The variable **INTRNT07** in the Tracker file indicates the status for the 2007 Internet Survey.

### ***2009 Internet Survey***

In 2009, contact letters were sent to 5613 HRS respondents, inviting them to participate in the 2009 Internet Survey. The 2009 Internet Survey questionnaire covered numerous topical areas including health (physical and mental; health behaviors), psychosocial questions, economics (income, assets, expectations, and consumption), and retirement.

The 2009 Internet field period was from March 2009 through August 2009. The 2009 Internet sample was drawn from respondents who reported internet access in the HRS 2008 core survey, plus those who did not respond to the 2008 core survey but had been selected for the 2003, 2006, or 2007 internet surveys. As was



the case in prior HRS internet surveys, roughly 20% of the eligible pool was reserved for a control group. A total of 4,433 respondents completed the 2009 Internet Survey, for a simple response rate of 77.2%. The variable **INTRNT09** in the Tracker file indicates the status for the 2009 Internet Survey.

### ***2011 Internet Survey***

In 2011, contact letters were sent to 5,742 HRS respondents, inviting them to participate in the 2011 Internet Survey. The 2011 Internet Survey questionnaire covered numerous topics including health, cognition, well-being, economics (income, assets, expectations, and consumption), and retirement.

The 2011 Internet field period was from May 2011 through September 2011. The 2011 internet sample was drawn from respondents who completed their core HRS interview on or before January 31, 2011 and reported internet access. The sample included respondents who were in the 2009 internet sample (and any new spouses) and who still had Internet access, along with an 80% random subsample of new cohort respondents (Middle Baby Boomer) and an 80% random subsample of panel respondents who reported internet access in 2010 but not in the prior wave (newly acquired Internet access). The remaining 20% of the latter two groups were assigned to the control group. A total of 4,590 respondents completed the 2011 Internet Survey, for a simple response rate of 79.9%. The variable **INTRNT11** in the Tracker file indicates the status for the 2009 Internet Survey.

### ***2013 Internet Survey***

In 2013, contact letters were sent to 7,739 HRS respondents, inviting them to participate in the 2013 Internet Survey. The 2013 Internet Survey questionnaire covered numerous topics including siblings, prescription drugs, cognition, economics (income, assets, expectations, and consumption), and residential history.

The 2013 Internet field period was from May through September 2013. The sample was drawn from respondents who completed a core 2012 HRS interview, and reported internet access. The sample included respondents who were in the 2011 internet sample (and any new spouses) who still had internet access, along with anyone who reported internet access in 2013, whether or not they were in the 2011 internet sample. A subsample of 20% was placed in the control group. A total of 5810 respondents completed the survey for a simple response rate of 75%. The variable **INTRNT13** in the Tracker file indicates the status for the 2009 Internet Survey.

### ***4L. NDI (NSCORE, NMONTH, NYEAR, xNDIFLAG: NO LONGER IN TRACKER)***

The HRS contract agreement with the providers of the National

Death Index has expired since the last release of the Tracker file, so relevant variables have been temporarily removed (NSCORE, NMONTH, NYEAR, and xNDIFLAG). We hope to update them in a subsequent release. Please see section 4M for alternative death date variables.

**4M. DEATH DATES AND LAST KNOWN ALIVE DATE (EXDEATHMO, EXDEATHYR, EXDODSOURCE, KNOWNDECEASEDMO, KNOWNDECEASEDYR, KNOWNDECEASEDSOURCE, LASTALIVEMO, LASTALIVEYR, LASTALIVESOURCE)**

The Tracker file contains the death date variables EXDEATHMO and EXDEATHYR, which are obtained from an interview, and EXDODSOURCE, which indicates the interview source for these variables.

In addition, we have created six new variables: KNOWNDECEASEDMO, KNOWNDECEASEDYR, KNOWNDECEASEDSOURCE, LASTALIVEMO, LASTALIVEYR, and LASTALIVESOURCE. The KNOWNDECEASED variables reflect our best information about death date for deceased respondents. These variables will match EXDEATHYR and EXDEATHMO for respondents with an exit interview where a death date was provided, or whose spouse or partner has provided a death date during a core interview subsequent to the respondent's death. For other cases, this date reflects either a death reported by someone with knowledge of the respondent, or a date imputed from the last date we knew the respondent was alive and the date on which we learned the respondent was deceased.

The new variables LASTALIVEMO, LASTALIVEYR and LASTALIVESOURCE represent the date on which we last knew the respondent was alive. This date comes from the most recent of: the respondent's last core interview, the last core interview of the respondent's spouse/partner during which the respondent was reported to be alive, or the last date during field activity on which we talked to the respondent or someone with direct knowledge of the respondent.

**4N. ADAMS Wave 1 (ADAMS1)**

The Aging, Demographics, and Memory Study (ADAMS) is a supplement to the Health and Retirement Study with the specific aim of conducting a population-based study of dementia. Data in ADAMS1 were collected between 2001 and 2005. The purpose was to gather additional information on respondents' cognitive status and assign a diagnosis of (1) dementia, (2) cognitive impairment but not demented (CIND), or (3) non-case. The data collected allow researchers to estimate the prevalence, predictors, and outcomes of dementia in the U. S. elderly population. A group of 1,770 HRS respondents, age 70 or older, was selected from the 2000 and 2002 waves based on the score on the self or proxy cognitive assessment measure. Among this group, assessments were completed for 856 respondents. The variable **ADAMS1** in the

Tracker data indicates the status for participation in this study.

#### **4O. Physical Measures (PHYMSR04)**

In 2004, HRS administered a set of physical measures to a random subsample of about 3,300 respondents. In 2006, this effort was expanded in terms of both the size of the sample covered and the measurements conducted, as HRS initiated what is referred to as the Enhanced Face-to-Face interview. The variable PHYMSR04 in the Tracker file indicates whether a respondent was selected to participate in the physical measures portion of the interview in 2004, and if selected, whether they completed the measures. For 2006 and beyond, the variable **EFTFASSIGN** (see 3D above) holds the respondents' permanent assignment for enhanced face-to-face rotation, which includes the Physical Measures component.

#### **4P. Biomarker Data Sets (GENETICS06, GENETICS08, GENETICS10, GENETICS12)**

These variables indicate whether a respondent is a member of the sample included in the Biomarker data sets of 2006, 2008, 2010 and 2012.

HRS collected blood-based biomarkers on half the sample in 2006, and the other half of the sample provided biomarker data in 2008. The first group was asked for blood samples again in 2010 and the second group gave repeat samples in 2012. The 2010 and 2012 samples each also include a random half of the new cohort enrolled in 2010.

For all years dried blood spot (DBS) samples have been assayed for these biomarkers: a). Total and HDL cholesterol, indicators of lipid levels b). Glycosylated hemoglobin (HbA1c) - an indicator of glycemic control over the past 2-3 months c). C-reactive protein (CRP), a general marker of systemic inflammation d). Cystatin C, an indicator of kidney functioning.

The blood tests were intended for all of those who were available for the Enhanced Face-to-Face Interview (EFTF) interview.

#### **4Q. Leave-Behind Questionnaires (DISAB04, PSYSOC04, PSYSOC06, PSYSOC08, PSYSOC10, PSYSOC12, PSYSOC14: NO LONGER IN TRACKER)**

The variables **DISAB04, PSYSOC04, PSYSOC06, PSYSOC08, PSYSOC10, PSYSOC12** and **PSYSOC14** have been removed from the Tracker file. For 2004, they are replaced by the variables JLBTTYPE and JV900 in Section LB of the core interview. For 2006 onward, they are replaced by xLBELIG and xLBCOMP in section LB.

Please see the [Leave Behind Questionnaire documentation report](#) for more information about the Psychosocial measures.

#### **4R. HEALTHY COGNITIVE AGING PROJECT (HCAP), 2016**

Sample indicators for the Healthy Cognitive Aging Project (HCAP16) are included in Tracker 2016. HCAP is a substudy within the HRS and is part of an ongoing international research collaboration to measure and understand dementia risk within ongoing longitudinal studies of aging around the world using a Harmonized Cognitive Assessment Protocol (also HCAP).

HCAP cases includes a pair of interviews, one by the Target Respondent (pre-selected HRS panel member, age 65+) and one by an Informant (nominated by the Target Respondent). Ideally, both interviews were conducted at the same time with a trained HRS interviewer in the Target Respondent's home.

The eligible sample consisted of pre-selected HRS Panel respondents, age 65 or older (birth year 1952 or earlier), who completed their 2016 interview before they become eligible for the HCAP study. Of those who were age-eligible and who completed their 2016 Core interview, those eligible for HCAP were selected via the following process:

- 1) Randomly select 1/2 of uncoupled Respondents.
- 2) Randomly select 1 Respondent from each coupled household.

While the majority of HCAP cases included both a Target Respondent and Informant interviews, there were some instances where a Target did not nominate any potential informants. There were also a few cases where only an informant interview was conducted as the Target Respondent was not well-enough to attempt an interview. HCAP was administered in both English and Spanish.

#### **4S. VENOUS BLOOD STUDY (VBS), 2016**

Sample indicators for the Venous Blood Study (VBS) 2016 are included in Tracker 2016. VBS 2016 is a sub study within the HRS that collected venous blood for the first time in the 2016 wave from panel respondents in a separate home visit by a trained phlebotomist in order to provide a substantially fuller picture of the health of a representative sample of older people. More information about VBS 2016 can be found [here](#) on the HRS public download site.

### **5. Wave-Specific Variables**

The wave-specific variables in the Tracker file may be divided into three groups: 1) those indicating a sample member's interview type, sample status, or study status in a given wave; 2) those indicating the relationships between the sample members in a household in a given wave; and 3) those describing other wave-specific information.

## **5A. Interview, Sample and Study Status**

### **5A1. *xIWTYPE*, *xIWWAVE*, and *xINSAMP***

In most of the HRS data waves, a sample member could fall into one of seven categories: (1) provided core interview; (2) eligible but did not provide core interview; (3) provided exit interview; (4) eligible but did not provide exit interview; (5) provided post-exit interview; (6) eligible but did not provide post-exit interview; and (7) not eligible in the wave. This information is captured in the variable *xIWTYPE*, which is the first variable a user might want to consult when checking a sample member's status in any given wave. By design, sample members in HRS 1992 or AHEAD 1993 can appear only in categories (1) and (2), and sample members in HRS 1994, AHEAD 1995 or HRS 1996 will not be in categories (5) and (6).

Closely related to *xIWTYPE* is *xIWWAVE*, a flag variable which collapses the interview type information into a single YES (1) or NO (0) dichotomy. *xIWWAVE* is equal to 1 if a sample member provided any type of interview; otherwise it is equal to 0 if the sample member was eligible but did not provide an interview. If a sample member was not eligible to be interviewed at a particular wave, *xIWWAVE* is blank.

While both *xIWTYPE* and *xIWWAVE* indicate whether an interview was obtained for a sample member in a given wave, *xINSAMP* indicates whether a sample member was supposed to be interviewed in the wave, and, if not, why. By design, some sample members should not be in certain data waves. Examples include (1) the (non-overlap) HRS members should not be in an AHEAD data wave, and vice versa; (2) the CODA or WB respondents should not be in any data wave prior to HRS 1998; (3) the EBB respondents should not be in any wave of data prior to HRS 2004, etc.

Tracker 2016 defines eight types of sample status: (1) in the sample; (4) members of HRS-AHEAD overlap households in which no one was interviewed at AHEAD wave 1, and who were dropped from the sample (i.e., were not contacted) at subsequent waves; (5) no longer in the sample because a complete exit (and, if necessary, a post-exit interview) has been obtained; (6) permanently dropped from the sample per request of the sample person, his/her spouse/partner, or other gatekeeper; (7) a deceased sample member for whom the field staff was unable to find an informant eligible to do an exit or post-exit interview at a previous wave; and (8) permanently dropped from the sample for any other reason. Which of these categories to include in non-response analyses will depend on the specific purpose, but in general, categories (4), (6), (7), and (8) should be regarded as eligible and therefore included in the denominator when calculating non-response and attrition rates.

The relationship among xIWTYPE, xIWWAVE, and xINSAMP is summarized below in Table 3.

**Table 3. The Relationship among xIWTYPE, xIWWAVE, and xINSAMP**

| <b>xIWTYPE</b>  | <b>xIWWAVE</b> | <b>xINSAMP</b>            |
|---|----------------|---------------------------|
| 1. Core interview provided<br>11. Exit interview provided<br>21. Post-exit interview provided             | 1. Yes         | 1. (Yes)                  |
| 5. Core interview NOT provided<br>15. Exit interview NOT provided<br>25. Post-exit interview NOT provided | 0. No          |                           |
| 99. Not in the sample this wave   | Blank.         | 2, 3, 4, 5, 6, 7, 8. (No) |

**5A2. xRESCODE**

Part of the HRS tracking procedure is to note reasons for non-interviews. For each sample member eligible to be interviewed in a wave, the variable xRESCODE indicates the final disposition of the case. Of particular importance is the distinction between a non-interview for the current wave only (xRESCODE codeframes between 5000 and 8999) and a permanent removal from the study (xRESCODE codeframes beginning with 9000).

Each sample member eligible to be interviewed in a wave will have a non-missing value on xRESCODE. The sample members of wave X with missing values on xRESCODE, therefore, will be those with a value other than 1 on xINSAMP.

**5A3. xSUBHH, and xSUBHHIW**

In HRS, sub-households are identified uniquely by the same household ID (HHID) AND sub-household ID (xSUBHH), and are usually regarded as the unit of household level analysis. An original household (xSUBHH = 0) may split off into a new xSUBHH for one of two reasons: dissolved marriage or partnership, or death of one of the partners. The sub-household ID for an eligible deceased sample member in a given wave is "3" or "4"; these values are never assigned to a sample member who is still alive. A respondent who is reported as deceased during the field period, but for whom we have not obtained an exit interview, will generally not be moved into xSUBHH of "3" or "4" until the exit interview is obtained, or until the surviving spouse /

partner reports the death in Section A of their own core interview.

xSUBHHIW describes whether a unique sub-household is interviewed in a given wave. A sub-household is considered interviewed in a wave (xSUBHHIW=1) if any of the members in the sub-household provided an interview in the wave, whether it was a core, exit or post-exit interview; and whether the interview was complete or partial. It is possible, therefore, for a sample member himself or herself not interviewed in a wave to have a value of 1 on xSUBHHIW. On the other hand, when no members in a sub-household were part of the sample in a wave (xINSAMP>1), xSUBHHIW will be missing for all the respondents in the sub-household.

## **5B. Inter-Respondent Relationship**

### **5B1. xCOUPLE, xCOUPID, xMARST, and xLIVARR**

When an eligible sample member in a given wave is coupled (married or living with a partner as if married), the variable **xCOUPLE** is equal to 1. Otherwise, it is 5. If one member of a couple is reported dead in a core or exit interview in wave x, the deceased respondent is assigned xSUBHH of "3" or "4", with xCOUPLE=1, while the surviving spouse or partner will be assigned xCOUPLE=5. (However, if the surviving spouse or partner has remarried or re-partnered, or was interviewed before the spouse / partner's death, xCOUPLE will equal 1).

Labeled as "xWave Household ID," **xCOUPID** is created to link the sample members who are married or partnered in a given wave. For sample members with xCOUPLE=1, xCOUPID is a concatenation of the person numbers of the sub-household members in ascending order. XCOUPID is also useful when a death has occurred. When one member of a couple is reported dead at wave x, and the partnership or marriage had not ended prior to the death, the deceased respondent is assigned xCOUPLE=1 and xCOUPID is established to describe the respondent's coupleness at death to facilitate linkage to the prior wave spouse or partner. The surviving spouse/partner is assigned 5 for xCOUPLE and xCOUPID is formed from his/her own PN followed by 000, unless he or she has formed a new couple.

**xMARST** indicates whether a respondent is married, divorced/separated, widowed, or has never been married. This variable is available beginning with HRS 2004 Core data. It is important to note that xMARST may be different from the marital status assigned in the core released data. The marital status variable in the Tracker file (xMARST) was constructed by examining several data sources. In cases of disagreement the marital status variable in the Core interview was not changed,

in order to preserve the flow through the instrument.

The basic procedure for generating xMARST is as follows. We first use core interview data to determine the marital status for those who provided core interviews at a given wave, thus creating xMARST. This variable is then modified to achieve cross-wave consistency based on one or more of the following: a) spouse/partner relationship, b) examination of cross sectional and longitudinal data, c) extrapolation based on the data, and sometimes d) common sense.

The "common sense" method stipulates that if a respondent was "never married" at wave t, then s/he was not married at any wave prior to t. Conversely, if a respondent was married, separated, or divorced at wave t, then s/he could not be "never married" at any wave after t. The "spouse/ partner relationship" method assumes that in a coupled household at wave t, if one respondent was married, the other should be married at the wave as well; and if one respondent was not married, the other should not be married either. The "extrapolation" method allows one to use a respondent's marital status at a given wave (t-1) to "approximate" his or her status at wave t when the latter information is not available.

**xLIVARR** describes a respondent's living arrangements and is available beginning with HRS 2002 Core data. For this variable, a respondent's marital status takes priority over any other living arrangements, so a married respondent who is separated and in another live-in relationship will be coded as "2 Married or partnered, not living with partner."

### **5B2. xFAMR and xFINR**

HRS respondents provide information at an individual level and, for some questions, at a household level. In particular, many questions about family or about income and wealth are asked of only one respondent in a household. The respondent providing family information in a given wave is called the Family Respondent (or Family R) in that wave, and the respondent providing financial data in a wave is called the Financial Respondent (or Financial R) in the wave. The family R and financial R designations are reflected in variables xFAMR and xFINR. For a single-person household, the respondent is designated to be both Family R and Financial R.

### **5B3. xNEWSP**

When an original sample member takes a new spouse or partner following his or her baseline interview, an attempt is made to interview the new person. New sample members entering in this way are indicated with xNEWSP=1 for the wave in which an interview was first attempted, and XNEWSP=0 for that person in



all other waves. There are no new spouses or partners in baseline waves. Thus, no sample members in HRS 1992 or AHEAD 1993 are considered new spouses or partners, and similarly, no sample members from the CODA, WB (1998), EBB (2004) and MBB (2010) entry cohorts are considered new spouses or partners.

#### **5B4. xPPN**

xPPN identifies the person number of the spouse or partner of a sample member if the sample member is part of a couple. For all the sample members still alive in a given wave, a concatenation of their own person numbers (PN) with xPPN in ascending order would generate xCOUPID. For deceased sample members, xPPN gives the person numbers of their spouses or partners in the last wave during which they were alive.

#### **5B5. xPROXY**

The HRS makes every effort to obtain core interviews directly with respondents themselves. In cases where the respondent is unavailable, unable, or unwilling to do the interview, a proxy is sought to provide an interview on behalf of the respondent (unless the respondent refuses to allow a proxy). Exit and post-exit interviews, on the other hand, are always conducted with proxies. The variable xPROXY indicates self-interviews, and classifies proxies into two types: spouse/partner, and any others.

### **5C. Other Wave-Specific Variables**

#### **5C1. xALIVE**

This variable contains information about vital status based on core and exit interviews and HRS tracking information. It provides four categories: (1) alive at the wave (xALIVE=1); (2) presumed alive as of the wave (xALIVE=2); (3) death reported in the wave (xALIVE=5); and (4) death reported in a prior wave (xALIVE=6).

If the respondent gave an interview in the wave or was contacted directly by an interviewer during the wave, or was reported to be alive by a spouse or partner, they are considered definitely alive (xALIVE=1). Similarly, anyone who met those criteria at a subsequent wave will be classified as xALIVE=1 at all prior waves. If contact of a less definitive type was made, and there was no report of death, we classify the respondent as presumed alive (xALIVE=2).

No National Death Index information is used in constructing xALIVE. Both sources of information should be used to classify vital status according to the goals of any specific analysis.

### **5C2. *xIWLANG*, *xIWMODE*, *xIWMONTH*, *xIWYEAR*, and *xAGE***

These variables indicate the language (English or Spanish) in which the interview was conducted, the mode (in-person or telephone), the month and year in which the interview began and the age the respondent was at a specific interview. If more than one birth date was provided in different waves, *xAGE* is constructed using a 'best source' method. A self-report was considered the most reliable, followed by a spouse report, and finally a non-spouse proxy report. The public data calculates age using the last date of birth given, regardless of the source. This can result in some discrepancies between *xAGE* and *BIRTHYR*. *xAGE* is calculated only for respondents who completed a core interview.

### **5C3. *xNURSHM***

*xNURSHM* indicates whether the respondent was reported to be in a nursing home or other health care facility at the time of the respondent's interview or the spouse/partner's interview. This can be reported either by the respondent or a proxy during the respondent's interview at question A028, or by a spouse or partner describing the respondent's whereabouts. This definition of nursing home is broader than the Census definition of nursing home as a category of institutional group quarters. The HRS definition includes some people living in other facilities such as assisted living facilities. Future waves of tracker will attempt to subdivide these categories based on additional information about the type of residence.

HRS follows respondents longitudinally into nursing homes or other facilities. By design, HRS does not attempt a baseline interview with potential new respondents who are nursing home residents. Given the young age at recruitment of recent cohorts, this policy had significant impact only for the 1993 baseline wave of AHEAD, where the decision was that interviews would not be attempted for nursing home residents, even if they were part of a couple. We did decide to include these AHEAD nursing home spouses or partners in subsequent waves, however, and attempted to interview them. Current policy is to attempt a baseline interview with a nursing home resident who is part of a couple whose other member is part of the non-institutionalized population.

### **5D. HRS Sample Weights**

An important change is underway for HRS sampling weights beginning with the early release weights for the HRS 2016 data. From 2000 to 2014, separate weights have been provided for HRS respondents who report living in nursing homes or other facilities, based on post-stratification to decennial Censuses

with updating for intercensal years. A careful review has determined that a significant fraction of HRS respondents who report at question A028 living in nursing homes or other facilities, live in facilities such as assisted living facilities, that are not counted by the Census as institutional group quarters. Therefore, they should not be counted as part of the Census nursing home population for post-stratification purposes.

A full correction requires making individual assignments of residence type corresponding to Census definitions before post-stratification by type of residence can be done. For the early release of 2016 we instead do a single post-stratification of all HRS respondents to the total population of the American Community Survey (ACS) without regard for residential type. This provides a more accurate distribution of weights than would a faulty two-sector post-stratification based on self-reported residence type matched to Census definitions. However, it means that the total population of HRS respondents who report living in nursing homes will be greater than the corresponding population in the ACS because it includes people who live in other types of facilities. Correcting that will require individual reassignments of residence type that we hope to conclude for the final release of 2016.

In order to note this change, new variables are created with a suffix, "E," on the names: xWGTHHE, xWGTRE, xWHYOHWTE, and xWHYORWTE. Section 5D2 describes this change in more detail.

**5D1. xWGTHH, xWGTR, xWGTRNH, xWHYOWGT, xWHYOHWT, and xWHYORWT**

xWGTHH is the sampling weight for analysis at the household level. xWGTR is the sampling weight for analysis at the respondent level; respondents included are those living in the community. xWGTRNH is the sampling weight for analysis at the respondent level; respondents included are those living in nursing homes. In this version of the Tracker, xWGTRNH is available for the 2000-2014 waves.

Since the HRS sample is a multi-stage probability sample of the United States, with oversamples of Blacks and Hispanics, unbiased estimates of population parameters require the use of these sampling weights. Through 2002, the household level and respondent level weights (xWGTHH and xWGTR) are post-stratified to the March Current Population Survey (CPS) for the year of data collection. Starting in 2004, the much larger ACS is used in place of the CPS for post-stratification purposes. Although the HRS samples only community dwellers at baseline, it follows respondents longitudinally into nursing homes and beginning in 2000 can represent that population reasonably well. The nursing

home sampling weight (xWGTRNH) is designed to post-stratify nursing home residents interviewed in a given wave (2000 to 2014) to represent the national population of nursing home residents by race, sex, and age in that year. Population totals by age and sex, and by race, are based on Census data in 2000 and 2010, with intercensal interpolations based on nursing home populations from the Minimum Data Set (MDS). Only nursing home residents who are age 55 or older and interviewed in the wave have a non-zero value for xWGTRNH.

The (community dwelling) respondent level weight xWGTR is zero when a respondent is not cohort eligible in wave x, is residing in a nursing home, or is deceased. There are also a very few cases with zero xWGTR because their birth cohort was not ascertained. The variable xWHY0WGT summarizes which of these conditions is responsible for a zero respondent level weight for cases with a core interview in wave x (available for waves 1992 through 2002). For the 2004 to 2014 waves, xWHY0HWT explains why a household has a zero weight and xWHY0RWT explains why a respondent has a zero weight.

Prior to 2004, HRS weights are generated based on WTCOHORT, not BIRTHYR. WTCOHORT is constructed using a "first-mention" or "entry-wave" birth year measure. This information may not be completely consistent with BIRTHYR, which takes into account corrections by the HRS staff to the birth year measure constructed in WTCOHORT. Starting in 2004, BIRTHYR has been used to generate the weights. Conversely, after considerable investigation and discussion with the NIA Data Monitoring Committee, race and ethnicity is now based exclusively on a "first-mention" basis. Any post-baseline reports of race are not used.

In the 2012 release of the Tracker file, the 2004-2010 weights were updated to incorporate several changes. First, the post-stratification method used to generate the household and respondent-level weights for the 2004 and later waves was modified to use ACS rather than CPS. Second, the updated weights are based on the first mention of race and ethnicity, reversing some changes that were made to the race and ethnicity variables in some previous versions of Tracker. Finally, non-original sample members who are no longer coupled with an original sample member are now assigned non-zero weights, so long as they are cohort eligible. As a result of these updates, the core household and respondent-level weights for the 2004-2010 waves will differ slightly from those in earlier releases of the Tracker file.

## **5D2. *xWGTHHE, xWGTR, xWHY0HWTE, xWHYORWTE***

ACS, the primary data source for HRS post-stratification since 2004, started counting residents of most group quarters, which include nursing homes, and providing person weights beginning in 2006. This allows HRS to use ACS to post-stratify both community residents and nursing home residents in one step. Therefore, starting 2016, HRS post-stratifies all respondents using the ACS 1-year Public Use Micro Sample (PUMS) data from the year corresponding to the survey year (e.g., ACS 2016 1-year data for HRS 2016). Therefore, nursing home residents who participated in a given wave no longer have zero weights under *xWGTR*. It should be noted that HRS samples households in the community only in the baseline of a given cohort.

Although ACS counts residents of nursing homes in the person population, it excludes nursing homes from households and does not provide weights for group quarter facilities. In order to keep HRS post-stratification consistent with ACS, households where the only member, or households where both members, were in nursing homes are assigned zero values for the household weight (*xWGTHHE*), as done prior to the 2016 wave. However, unlike prior waves, nursing home residents are assigned non-zero values for the respondent weight (*xWGTR*). This also means that while all categories under *xWHY0HWT* variable are applicable, categories under *xWHYORWT* variables corresponding to nursing home cases are no longer applicable. To note this change, new variables, *xWHY0HWTE* and *xWHYORWTE*, were created.

HRS post-stratification is done at the household level as well as the respondent level using ACS PUMS. At the household level, population totals of household cohort, race/ethnicity of household heads, coupleness status, household type and counts of financial units are used. Respondent level post-stratification controlled for birth cohort, gender, marital status and race/ethnicity.

## **5D3. *Sample Weights for 2004 Physical Measure and Leave-Behind Questionnaire (JPMSELWT, JPMWGTR, JWGTR\_PS AND JWGTR\_DB)***

Sample weights were developed for the 2004 Physical Measures and Leave-Behind components to account for differential probabilities of selection into the subsample and non-response to the specific component. These variables include:

### *Physical Measures:*

PHYMSR04- sample selection indicator for 2004 physical measures  
JPMSELWT - selection weight for the 2004 physical measures subsample

JPMWGTR - final respondent weight for 2004 physical measures subsample

*Psychosocial Leave-Behind:*

JWGTR\_PS - final respondent weight for 2004 psychosocial subsample

*Disability Leave-Behind:*

JWGTR\_DB - final respondent weight for 2004 disability subsample

A general description of the physical measures and leave-behind components can be found in the Data Description for the 2004 Core file, and the questionnaires for each are included under "Other" Content Area for the 2004 wave on the Questionnaire page on the HRS website. More detailed information on the selection indicators and sample weights is provided in the following documents:

[Physical Measures](#)

[Leave-Behind Questionnaires](#)

#### **5D4. Sample Weights for the 2006 Physical Measures, Biomarkers, and Psychosocial Leave Behind Questionnaire (KPMWGTR, KBIOWGTR and KLBWGTR)**

Sample weights were developed for the physical measure and biomarker components separately. Respondents who completed at least one physical measure were assigned a physical measures weight and respondents with at least one valid biomarker result were assigned a biomarker weight. The weights were adjusted for the differential probabilities of participation by dividing the HRS 2006 sample weight by the predicted probability of response to each component. The resulting interim weight was trimmed at the 5<sup>th</sup> and 95<sup>th</sup> percentiles and was then post stratified back to the entire 2006 HRS sample by age, sex, and race/ethnicity. The sample weights for each component are provided in the Tracker file. These include:

EFTFASSIGN - enhanced face-to-face sample indicator

KPMWGTR - final respondent weight for 2006 physical measures subsample

KBIOWGTR - final respondent weight for 2006 biomarker subsample

More information on the physical measures component including details on sampling, consent, administration, and rationale is provided in the following [User Guide](#).

Information on the biomarker collection can be found in the [Data Description for the 2006 Biomarker Data file](#).

Separate respondent level survey weights were constructed to adjust for non-response to the Leave Behind component. The HRS 2006 sample weight was adjusted for non-response by dividing it by the predicted probability of response to the questionnaire. Age, sex, race/ethnicity, marital status, health status, depressive symptoms, low cognition, IADL/ADL difficulties, mobility, poor vision, work status, and religious attendance were used in the non-response model. The resulting interim weight was trimmed at the 5<sup>th</sup> and 95<sup>th</sup> percentiles and was then post stratified back to the entire 2006 HRS sample by age, sex, and race/ethnicity. The sample weights, sample selection and eligibility indicators for the psychosocial component include the following variables:

EFTFASSIGN - enhanced face-to-face sample indicator  
KLBWGTR - final respondent weight for the 2006 leave behind questionnaire subsample

More information about the questionnaire can be found in the [User Guide](#).

#### **5D5. Sample Weights for the 2008-2014 Physical Measures, Biomarkers, and Psychosocial Leave-Behind Questionnaire**

Sample weights for the 2008-2012 Physical Measures, Biomarker, and Psychosocial Leave-Behind Questionnaire were developed using the same methodology as in 2006. The sample weights for each component are provided in the Tracker file. These include:

EFTFASSIGN - enhanced face-to-face sample indicator  
LPMWGTR - final respondent weight for the 2008 physical measures subsample  
LBIOWGTR - final respondent weight for the 2008 biomarker subsample  
LLBWGTR - final respondent weight for the 2008 leave behind questionnaire subsample

MPMWGTR - final respondent weight for the 2010 physical measures subsample  
MBIOWGTR - final respondent weight for the 2010 biomarker subsample  
MLBWGTR - final respondent weight for the 2010 leave behind questionnaire subsample

NPMWGTR - final respondent weight for the 2012 physical measures subsample  
NBIOWGTR - final respondent weight for the 2012 biomarker subsample  
NLBWGTR - final respondent weight for the 2012 leave behind questionnaire subsample

OPMWGTR - final respondent weight for the 2014 physical measures subsample  
OBIOWGTR - final respondent weight for the 2014 biomarker subsample  
OLBWGTR - final respondent weight for the 2014 leave behind questionnaire subsample

## **6. Important Data Changes in Early Tracker 2016**

### **6A. NDI Variables Removed**

Please see [Section 4L](#).

### **6B. New Variables**

Please see Sections [4J](#), [4M](#), [4R](#) and [4S](#) for a description of new variables.

### **6C. Two lines removed**

It was discovered that two lines in the Tracker file were spawned in error (as described below) during the 1996 data collection wave, and these lines were removed.

HHID 056354, PN 031: This line was a duplicate of HHID 056354 PN 020. The 031 line was spawned in error in the 1996 data collection wave. It has been removed for this Tracker release, and data collected on that line was transferred to the 020 line. The variables on 030's line pertaining to coupled status were updated where applicable to reflect only 020 as a partner.

HHID 084287, PN 011: This line was spawned in error in the 1996 data collection wave. It has been removed for this Tracker release and the 010 line has been updated to reflect the fact that it was a single respondent household.

## **7. Obtaining the Data**

### **7A. Registration and Downloading the Data**

HRS data are available for free to researchers and analysts at the HRS Website. In order to obtain public release data, you must first register at our [Website](#). 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.



## **7B. 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:

- not attempt to identify respondents.
- not transfer data to third parties except as specified.
- not share your username and password.
- include specified citations in work based on HRS data.
- provide information to us about publications based on HRS data.
- report apparent errors in the HRS data or documentation files.
- 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 Statement](#)" at the [Public File Download Area](#) of the [HRS Website](#).

## **7C. 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  
P.O. Box 1248  
Ann Arbor, MI (USA) 48106-1248

You may also contact us by e-mail at [hrsquestions@umich.edu](mailto:hrsquestions@umich.edu) with "Attn: Papers and Publications" in the subject line.

## **8. If You Need to Know More**

This document is intended to serve as a brief overview and to provide guidelines for using the Tracker 2016 data. If you have questions or concerns that are not adequately covered here or on our [Website](#), or if you have any comments, please contact us. We will do our best to provide answers.

### **8A. 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, please visit the HRS [Website](#).

**8B. Contact Information**

Internet: Help Desk at our Website

E-mail: [hrsquestions@umich.edu](mailto:hrsquestions@umich.edu)

Postal service:

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

FAX: (734) 647-1186

## APPENDIX

### 9. Merging the Tracker with other HRS Data

When merging the Tracker data with other HRS data, you should remember that the Tracker contains more records than other data releases for any given wave. The first step is, therefore, to subset the records from the Tracker that are present in the other data file. `xIWTYPE` provides necessary information for performing this task. In addition, since the Tracker contains only one record per sample member, the four respondents with overlap identifiers (as described in 3C Overlaps) will not match with other HRS data without some manipulations before merging. In fact, dealing with the respondents with multiple identities across waves proves to be the most complicated task when integrating the Tracker data with other HRS data.

In what follows, we provide some coding examples for merging the Tracker with various HRS data, covering most of the issues that users may encounter when performing similar merges of their own. For simplicity, we assume that all the data, including the Tracker and other HRS data, are stored in the same directory (or SAS data library).

#### 9A. HRS 1992 Merges

Merging the Tracker with the HRS 1992 data is complex because of the HRS-AHEAD overlaps explained in Section 3C. Two different strategies are available for working with HRS 1992 data. First, one may consider the entire, original HRS sample as the data source of the analysis. In this strategy, one wants to retain all the data records as released in the HRS 1992 Core data, that is,  $N=12,652$ . Alternatively, one may consider only the non-overlap cases as the HRS respondents, leaving the HRS overlaps to AHEAD. In this strategy, the information in the HRS 1992 Core data for the overlap cases will be dropped, and  $N=12,521$ .

While we have no intention to force users to take one strategy over the other, it is suggested that the second strategy be used whenever the HRS 1992 data are involved in a longitudinal analysis. The overlap cases are no longer treated as part of the HRS sample in the later waves.

The following examples illustrate these two strategies when merging with a dataset called HEALTH

##### 9A1. Retaining the Entire, Original HRS Sample

```
(i) SAS Code
    data tracker;
    set dat.trk2016;
    if ovresult=1 then do; /* Convert AHEAD IDs into Original HRS
IDs */
```

```

        hhid=ovhhid;
        pn=ovpn;
    end;
run;

proc sort data=tracker;
    by hhid pn;
run;

proc sort data=dat.health out=health(keep=hhid pn v1 );
    by hhid pn;
run;

data healtht;
    merge tracker health(in=inh);
    by hhid pn;
    if inh;
run;

```

(ii) Stata Code

```

use trk2016
replace HHID=OVHHID if OVRESULT==1
replace PN=OVPN if OVRESULT==1
sort HHID PN
save tracker, replace

use health
sort HHID PN

merge HHID PN using tracker
keep if _m==3
save healtht, replace

```

(iii) SPSS Code

```

GET FILE "c:/temp/health.sav".
SORT CASES BY hhid pn.
execute.
save outfile='C:/temp/health2.sav'.

GET FILE "C:\temp\trk2016.sav".
* Convert AHEAD IDs into Original HRS IDs.
if (ovresult =1)hhid=ovhhid.
if (ovresult =1)pn=ovpn.
SORT CASES BY hhid pn.
execute.
save outfile='C:/temp/tracker.sav'.

MATCH FILES /FILE="C:/temp/health2.sav"
/IN=h

```

```

    /FILE='C:\temp\tracker.sav'
    /BY hhid pn
    select if h=1.
    EXECUTE.

```

```

save outfile='C:/temp/healtht.sav'.

```

## **9A2. Keeping Non-Overlaps Only**

(i) SAS Code

```

proc sort data=dat.trk2016 out=tracker;
  by hhid pn;
run;

```

```

proc sort data=dat.health out=health(keep=hhid pn v1);
  by hhid pn;
run;

```

```

data healtht;
  merge tracker(in=int) health(in=inh);
  by hhid pn;
  if ovresult=0 and inh;          /* Keep only Non-Overlap
Cases */
run;

```

(ii) Stata Code

```

use trk2016
sort HHID PN
save tracker, replace

```

```

use health
sort HHID PN

```

```

merge HHID PN using tracker
keep if _m==3 & OVRESULT==0    /* Keep Only Non-Overlap
Cases */
save healtht, replace

```

(iii) SPSS Code

```

GET FILE "c:/temp/health.sav".
SORT CASES BY hhid pn.
execute.
save outfile='C:/temp/health3.sav'.

```

```

GET FILE "C:\temp\trk2016.sav".
SORT CASES BY hhid pn.
execute.
save outfile='C:/temp/tracker2.sav'.

```

```

MATCH FILES /FILE="C:/temp/health3.sav"
  /IN=h
  /FILE='C:\temp\tracker2.sav'
  /BY hhid pn.
select if h=1 and ovresult=0.
EXECUTE.

save outfile='c:/temp/healtht.sav'.

```

### **9B. HRS 1998 Merges**

In this example, we merge the tracker file with dataset h98a\_r. In this case, there is one respondent identified in the 1998 data by their OVHHID and OVPN, so we revert the identifiers in tracker to match the released data before merging.

#### (i) SAS Code

```

data tracker;
  set dat.trk2016;
  if ovresult=8 and hhid eq "205864" then hhid=ovhhid;
  if ovresult=8 and hhid eq "205864" then pn=ovpn;
  /* Convert AHEAD Inter-Respondent Overlap ID into Original */
  run;

proc sort data=tracker;
  by hhid pn;
  run;

proc sort data=dat.h98a_r out=h98a_r (keep=hhid pn f968 f970);
  by hhid pn;
  run;

data h98a_rt;
  merge tracker h98a_r(in=inh);
  by hhid pn;
  if inh;
  run;

```

#### (ii) Stata Code

```

use trk2016
replace HHID=OVHHID if OVRESULT==8 & HHID~="205864"
replace PN=OVPN if OVRESULT==8 & HHID~="205864"

sort HHID PN
save track, replace

use h98a_r
keep HHID PN F968 F970

```

```

sort HHID PN

merge HHID PN using track
keep if _m==3
save h98a_rt, replace

```

(iii) SPSS Code

```

GET FILE "C:\temp\trk2016.sav".

* Convert AHEAD Inter-Respondent Overlap ID to original.
  if (ovresult =8) and (HHID <> '205864') hhid=ovhhid.
if (ovresult =8) and (HHID <> '205864') pn=ovpn.
execute.
SORT CASES BY hhid pn.
execute.
save outfile='C:/temp/tracker.sav'.

GET FILE "C:\temp\h98a_r.sav"/keep hhid pn f968 f970.
SORT CASES BY hhid.
execute.
save outfile='C:\temp\h98a_r2.sav'.

MATCH FILES /FILE="C:/temp/h98a_r2.sav"
  /IN=h
  /FILE='C:\temp\tracker.sav'
  /BY hhid pn.
select if h=1.
EXECUTE.
save outfile='C:/temp/h98a_rt.sav'.

```

### **9C. HRS 2010 Merges for 2012 Overlaps**

As described in Section 3C, there are two respondents who appear in 2010 released data under their OVHHID and OVPN. In this example, we convert their identities in the core 2010 data to match both tracker and the 2012 data release.

(i) SAS Code

```

data tracker;
  set dat.trk2016;
  run;

proc sort data=tracker;
  by hhid pn;
  run;

data h10a_r;
  set dat.h10a_r;

```

```
if hhid = '526934' and pn = '010' then do;
    hhid = '525520'; pn = '020';
end;

else if hhid = '529766' and pn = '010' then do;
    hhid = '520845'; pn = '020';
end;

run;

proc sort data=h10a_r (keep=hhid pnma019);
    by hhid pn;
run;

data h10a_rt;
    merge tracker h10a_r(in=inh);
    by hhid pn;
    if inh;
run;
```