

HEALTH AND RETIREMENT STUDY

Tracker 2012
Early, Version 1.0
June, 2014

Data Description

TABLE OF CONTENTS

TABLE OF CONTENTS.....	III
DATA DESCRIPTION AND USAGE	1
1. INTRODUCTION.....	1
2. THE STRUCTURE OF TRACKER 2012	1
2A. VARIABLE LISTING AND DESCRIPTION.....	1
2B. NAMING CONVENTION FOR THE WAVE-SPECIFIC VARIABLES.....	4
2C. CHARACTER VERSUS NUMERIC	4
3. PERMANENT CHARACTERISTICS.....	5
3A. DEMOGRAPHIC INFORMATION.....	5
3B. FIRST INTERVIEW.....	5
3C. OVERLAPS.....	5
3D. ENHANCED FACE-TO-FACE ASSIGNMENTS	6
3E. ADDITIONAL VARIABLES WITH PERMANENT CHARACTERISTICS	6
4. ANCILLARY STUDIES AND SUPPLEMENTAL VARIABLES	6
4A. 1999 PILOT MAIL SURVEY (MAIL99)	6
4B. 2001 MAIL SURVEYS (CAMS01 AND HUMS01)	6
4C. 2003 MAIL SURVEYS (CAMS03 AND DIAB03)	6
4D. 2005 MAIL SURVEYS (CAMS05 AND PDS05).....	7
4E. 2007 MAIL SURVEYS (CAMS07, DVSA07 AND DVSB07).....	8
4F. 2009 MAIL SURVEY (CAMS09)	9
4G. 2011 MAIL SURVEY (CAMS11, HCMS11).....	9
4H. INTERNET SURVEYS (INTRNT03, INTRNT06, INTRNT07, INTRNT09 AND INTRNT11).....	10
4I. NDI (NSCORE, NMONTH, NYEAR, DEATHMO, DEATHYR AND DODSOURCE)	12
4J. ADAMS WAVE 1 (ADAMS1)	12
4K. PHYSICAL MEASURES (PHYMSR04)	13
4L. LEAVE-BEHIND QUESTIONNAIRE (DISAB04, PSYSOC04, PSYSOC06 AND PSYSOC08)	13
4M. VETERANS SURVEY.....	14
5. WAVE-SPECIFIC VARIABLES	14
5A. INTERVIEW, SAMPLE OR STUDY STATUS.....	14
5A1. <i>xiWTYPE</i> , <i>xiWWAVE</i> , and <i>xINSAMP</i>	14
5A2. <i>xRESCODE</i>	15
5A3. <i>xSUBHH</i> , and <i>xSUBHHIW</i>	15
5B. INTER-RESPONDENT RELATIONSHIP	16
5B1. <i>xCUPLE</i> , <i>xCOUPID</i> , <i>xCORES</i> , and <i>xMARST</i>	16
5B2. <i>xFAMR</i> and <i>xFINR</i>	17
5B3. <i>xNEWSP</i>	17
5B4. <i>xPPN</i>	18
5B5. <i>xPROXY</i>	18
5C. OTHER WAVE-SPECIFIC VARIABLES.....	18
5C1. <i>xALIVE</i>	18
5C2. <i>xiWLANG</i> , <i>xiWMODE</i> , <i>xiWMONTH</i> , <i>xiWYEAR</i> , and <i>xAGE</i>	18
5C3. <i>xNURSHM</i>	18
5C4. <i>Changes to GNURSHM, HNURSHM, JNURSHM, KNURSHM and LNURSHM</i>	19
5D. HRS SAMPLE WEIGHTS.....	21

5D1. <i>xWGTHH, xWGTR, xWGTRNH, xWHYOWGT, xWHYOHWT, and xWHYORWT</i>	21
5D2. <i>Changes to the HRS Sample Weights</i>	22
5D3. <i>Sample Weights for 2004 Physical Measure and Leave-Behind Questionnaire (JPMSELWT, JPMWGTR, JWGTR_PS AND JWGTR_DB)</i>	35
5D4. <i>Sample Weights for the 2006 Physical Measures, Biomarkers, and Psychosocial Leave Behind Questionnaire (KPMWGTR, KBIOWGTR and KLBWGTR)</i>	36
5D5. <i>Sample Weights for the 2008 Physical Measures, Biomarkers, and Psychosocial Leave Behind Questionnaire</i>	37
6. IMPORTANT DATA CHANGES IN EARLY VERSION 1.0 OF TRACKER 2012	37
6A. <i>RACE, HISPANIC, AND GENDER UPDATED</i>	37
7. OBTAINING THE DATA	37
7A. <i>REGISTRATION AND DOWNLOADING THE DATA</i>	37
7B. <i>CONDITIONS OF USE</i>	37
7C. <i>PUBLICATIONS BASED ON DATA</i>	38
8. IF YOU NEED TO KNOW MORE	38
8A. <i>HRS INTERNET SITE</i>	38
8B. <i>CONTACT INFORMATION</i>	38
APPENDIX	I
9. MERGING THE TRACKER WITH OTHER HRS DATA	I
9A. <i>HRS 1992 MERGES</i>	I
9A1. <i>Retaining the Entire, Original HRS Sample</i>	i
9A2. <i>Keeping Non-Overlaps Only</i>	ii
9B. <i>HRS 1998 MERGES</i>	III

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 2012, Early Version 1.0) covers all types of interviews (core, exit, and post-exit) and contains information for panel cases (AHEAD, HRS, CODA, WB, and EBB) through HRS 2012. Structurally, Tracker 2012 Early Version 1.0 retains all the information contained in Tracker 2010 Final, Version 1.0 and adds information from the 2012 Core release.

This document, complementary to the Tracker file codebook, consists of nine sections. Section 2 describes the basic structure of Tracker 2012. 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 further information. 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 hqsquestions@umich.edu.

2. The Structure of Tracker 2012

2A. Variable Listing and Description

There are 397 Variables in Tracker 2012. These are listed in Table 1 according to the order in which they appear in the Tracker data and the codebook. First on the list are two 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. These variables either provide some basic demographic information for a respondent, or describe his or her 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.)?" In the 2012 Tracker file (Early, Version 1.0), the 2012 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 constants are several supplemental variables that are related to the National Death Index (NDI) and the HRS

Ancillary studies, including Off-Year Mail Surveys, Internet surveys, leave-behind questionnaires and ADAMS.

The final group includes twelve sets of wave-specific variables, with one set representing each of the twelve HRS survey waves to date. The waves are differentiated by the leading letter in variable names, from A to N (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 2012 Tracker (Early, Version 1.0): Listing and Description

Name	Type	Length	Label
Group A. Primary Identifiers and Permanent Characteristics			
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
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			
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
DEATHMO	Num	2	Month of Death

DEATHYR	Num	4	Year of Death
DIAB03	Num	2	2003 Diabetes Sample Status
DISAB04	Num	2	Sample Status for Disability Qnaire
DODSOURCE	Num	1	Source of Exit Date of Death
DVSA07	Num	2	2007 DVS-A Sample Status
DVSB07	Num	2	2007 DVS-B Sample Status
GENETICS06	Num	1	Member of 2006 GWAS Sample
GENETICS08	Num	1	Member of 2008 GWAS Sample
GENETICS10	Num	1	Member of 2010 GWAS Sample
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
MAIL99	Num	2	1999 Mailout Pilot Survey Status
NMONTH	Num	2	NDI Month of Death
NSCORE	Num	3	NDI Match Score
NYEAR	Num	4	NDI Year of Death
PDS05	Num	2	2005 PDS Sample Status
PHYMSR04	Num	2	Sample Status for Physical Measures
PSYSOC04	Num	2	Sample Status for PsychSocial Qnaire 2004
PSYSOC06	Num	2	Sample Status for PsychSocial Qnaire 2006
PSYSOC08	Num	2	Sample Status for PsychSocial Qnaire 2008
VA13	Num	2	2013 VA Sample Status

Group C. Wave-Specific Variables

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
xCORES	Num	1	Wave X Co-residence Status
xCOUPID	Char	6	Wave X Wave Household ID
xCUPLE	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
xiLBWGTR	Num	5	Respondent Weight for the Wave X Leave Behind Qnaire
xMARST	Num	1	Wave X Marital Status
xNDIFLAG	Num	1	NDI Wave X Alive or Deceased Flag
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 Physical 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
xWGTR	Num	8	Wave X Weight: Respondent Level
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
xWHY0HWT	Num	1	Wave X Why Zero Household Level Weight
xWHY0RWT	Num	1	Wave X Why Zero Respondent Level Weight
xWHY0WGT	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, B, C, D, E, F, G, H, J, K, L, M and N are used as indicators of the HRS waves. Each wave-specific variable is named with a leading wave indicator according to the letter listed in Table 2 (below).

All of the wave-specific variables share the same naming structures across waves. For example, ANEWSP, BNEWSP, CNEWSP, DNEWSP, ENEWSP, FNEWSP, GNEWSP, HNEWSP, JNEWSP, KNEWSP, LNEWSP, MNEWSP and NNEWSP are, respectively, new spouse or partner flags for HRS 1992, AHEAD 1993, HRS 1994, AHEAD 1995, HRS 1996, HRS 1998, HRS 2000, HRS 2002, HRS 2004, HRS 2006, 2008 HRS 2010 and HRS 2012. 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 2012 will be named as NNEWSP.)

Table 2. HRS Data Waves and Wave Indicators

HRS Data Wave	Wave Indicator
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

2C. Character Versus Numeric

Following a protocol used in all of the recent HRS data releases, the ID variables are stored in character format. In Tracker 2012, the ID variables include 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, that is, 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. Please also see Section 6A for changes made to the Race, Gender and Hispanic variables.

3B. First Interview

The variable FIRSTIW contains the information on the first time a sample member actually 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 2012, there are 616 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 OVHHID and OVPN.

There are two basic 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 sub-sampling 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, 109 were actually interviewed in AHEAD, and 25 were not.

Second, there are two cases (as of this version of the Tracker) belonging to what we call "household merge overlap" cases which result from intermarriage among respondents who entered the study in separate households. In one case, an AHEAD married 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. In the other, 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.

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

Starting in 2004, HRS began administering physical measures and biomarker tests for our respondents. The sample is structured so that half our respondents receive the enhanced interview each wave. The variable EFTFASSIGN holds the respondents' permanent assignment for enhanced face-to-face rotation.

3E. Additional Variables with Permanent Characteristics

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

4. Ancillary Studies and Supplemental Variables

4A. 1999 Pilot Mail Survey (MAIL99)

In 1999, 2001, 2003, 2005, 2007 and 2009, questionnaires were mailed to subsamples of the HRS. 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 Mail Survey consisted of two different questionnaires mailed to separate sub samples of HRS respondents. One version of the 2001 Mail Survey asked respondents about their consumption (household expenditures), time spent on activities, and prescription drugs. This survey is referred to as the 2001 Consumption and Activities Mail Survey (CAMS). The HRS data file [CAMS01_R] and supporting documentation can be found on the HRS web site. The variable CAMS01 indicates whether a respondent was selected to receive a 2001 CAMS questionnaire, and if selected, whether the questionnaire was returned. The second version of the 2001 Mail Survey, the Human Capital Mail Survey (HUMS) asked respondents about college expenses (human capital investments) associated with children who had attended college. The HRS data file [HUMS01_C] and supporting documentation can be found on the HRS web site. 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 CAMS Mail Survey asked respondents about their consumption (household expenditures), and time spent on activities. The pool for this sub sample 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 Consumption and Activities Mail Survey, and if selected, whether the questionnaire was returned. There was no Human Capital Mail Survey (HUMS) in 2003.

The HRS 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 Consumption and Activities Mail Survey (CAMS). This exclusion is random, with the slight exception that proxy cases from 2000 were ineligible for CAMS but eligible for the Diabetes Study, 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 (n=2,514), and if selected, whether the questionnaire was returned (n=4,415) or had died prior to the start of the study (n=129). The variable is further broken down into respondents who returned a questionnaire with the HbA1c blood kit (n=1,233) and those who did not return the blood kit (n=668).

4D. 2005 Mail Surveys (CAMS05 and PDS05)

The 2005 CAMS Mail Survey asked respondents about their consumption (household expenditures), and time spent on activities. The pool for the sub sample 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 CAMS but not eligible for the other 2005 Mail Survey (Prescription Drug Study (PDS05)), they were selected for the 2005 CAMS. If the respondent was eligible for CAMS and also eligible for the other 2005 mail sub sample (PDS05), 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 CAMS and also eligible for the other 2005 mail sub sample (PDS05), and they were an uncoupled household, half of the eligible respondents were selected randomly for CAMS 2005, and half for the other 2005 mail sub sample (PDS05). 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 sub sample of HRS respondents who were not in some other mail survey conducted in 2005.

In coupled households, two versions of the CAMS05 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 Consumption and Activities Mail Survey (n=8124), and if selected, whether the questionnaire was returned (n=5,815). 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 (PDS05) is the first of a two-wave mail survey designed to track changes in prescription drug utilization as Medicare Part D, the prescription drug benefit, is 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 is comprised of HRS respondents born in 1942 or earlier (65th 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 is expected to have its greatest impact on persons without drug coverage, and persons of low income and 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 are 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 (n=5,654), and if selected, whether the questionnaire was returned or a phone interview was completed (n=4,684). The variable is further broken down into respondents who returned a questionnaire with the medication list (n=4,376) and those who did not return the medication list (n=308).

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

The 2007 CAMS Mail Survey asked respondents about their consumption (household expenditures), and time spent on activities. The pool for this sub sample 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 Consumption and Activities Mail Survey (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 DVS Mail Survey, 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).

4F. 2009 Mail Survey (CAMS09)

The 2009 CAMS 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 Consumption and Activities Mail Survey (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.

4G. 2011 Mail Survey (CAMS11, HCMS11)

The 2011 CAMS 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 sub-sample 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 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 CAMS data file contains data from 6,531 respondents of which 1,613 are new cohort cases.

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 Consumption and Activities Mail Survey (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 and 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.

4H. Internet Surveys (INTRNT03, INTRNT06, INTRNT07, INTRNT09 and INTRNT11)

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 sub sample 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 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 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%.

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 2013 through September 2013. The 2013 Internet 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, irrespective of whether they were in 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%.

4I. NDI (NSCORE, NMONTH, NYEAR, DEATHMO, DEATHYR AND DODSOURCE)

In addition to the information on vital status obtained by HRS through tracking of respondents (see xALIVE variables), the HRS seeks matches to the National Death Index for persons who are reported as deceased or who are not known to be alive through contact during tracking. The Tracker file contains information derived from finder files submitted to the National Center for Health Statistics (NCHS), in 1995, 2000, 2002, 2004, 2006, and 2008. For all submitted cases that were flagged as valid by NCHS and verified by HRS staff, the Tracker file contains year and month of death, match score, and an alive/deceased flag (NYEAR, NMONTH, NSCORE, xNDIFLAG). The NDI month and year of death may differ from DEATHMO and DEATHYR, which are obtained from an interview; DODSOURCE indicates the interview source for month and year of death variables.

4J. 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.

4K. 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, which 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 will repeat 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). Results for C-reactive protein and cystatin C are forthcoming. The variable PHYMSR04 in the Tracker file indicate 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. The variable EFTFASIGN holds the respondents' permanent assignment for enhanced face-to-face rotation for 2006 and beyond, which includes the Physical Measures component.

4L. Leave-Behind Questionnaire (DISAB04, PSYSOC04, PSYSOC06 and PSYSOC08)

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

In 2006 and 2008, the Leave-Behind Questionnaire was incorporated into the Enhanced Face-to-Face Interview and covered half of the full sample in 2006, and the other half in 2008. From 2008 on, every household will repeat the enhanced face-to-face interview, including the Leave-Behind Questionnaire, every other wave. In 2006 and 2008 only the Participant Lifestyle or Psychosocial Questionnaire was administered. In the Tracker file, the variable DISAB04 indicates whether a respondent was selected to participate in the 2004 Disability Leave-Behind Questionnaire and if selected, whether

the questionnaire was returned. The variables PSYSOC04, PSYSOC06 and PSYSOC08 indicate whether a respondent was selected to participate in the Psychosocial Leave-Behind Questionnaire for a given wave, and if selected, whether the questionnaire was returned.

4M. Veterans Survey

In June 2013, questionnaires were mailed to 3811 HRS respondents who had reported military service in the past. The 2013 Veterans Mail Survey field period was June 2013 through September 2013. The 2013 Veterans questionnaire covered numerous topics including general health, use of VA services, their experience in the military, and friends from the military. The data file for the 2013 Veterans Mail Survey contains data for 1,871 respondents. The simple response rate for the 2013 Veterans Mail Survey is 49 percent.

5. Wave-Specific Variables

Roughly speaking, the wave-specific variables in Tracker 2012 may be divided into three groups: those indicating a sample member's interview, sample, or study status in a given wave; those indicating the relationships between the sample members in a household in a given wave; and those describing other wave-specific information.

5A. Interview, Sample or 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. For 2010 only, we have added the category 'completed interview. This information is captured in variable *xIWTYPE*, 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, and (3) the EBB respondents should not be in any wave of data prior to HRS 2004.

Tracker 2012 defines eight types of sample status: (1) in the sample; (2) belonging to an entry cohort not interviewed this wave; (3) not yet entered, but belonging to the cohort interviewed in this wave; (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, xIWAVE, and xINSAMP may be better seen in Table 3, where the concepts "in the wave" and "in the sample" are mathematically delineated.

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 and a permanent removal from the study. Each can generate one of several possible result codes indicating a specific reason.

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 of 2, 3, 4, 5, 6, 7, 8 or 9 on variable xINSAMP.

5A3. xSUBHH, and xSUBHHIW

In HRS, sub-households, which are split from original households and identified uniquely by the same household ID (HHID) AND sub-household ID (xSUBHH), are usually regarded as the unit of household level analysis. An original household may split off primarily for one of the two reasons: dissolved marriage or partnership, or death in the household. The sub-household ID for an eligible deceased sample member in a given wave is normally "3" or "4"; these values are never assigned to any sample member still alive.

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.

Table 3. The Relationship among xIWTYPE, xIWAVE, and xINSAMP

xIWTYPE	xIWAVE	xINSAMP

Core Interview Provided (1)		
Exit Interview Provided (11)	Yes (1)	
Post-Exit Interview Provided (21)		
-----	-----	Yes (1)
Core Interview Not Provided (5)		
Exit Interview Not Provided (15)	No (0)	
Post-Exit Interview Not Provided (25)		
-----		-----
Not in the Wave (99)		No (2,3,4,5,6,7,8)
-----		-----

5B. Inter-Respondent Relationship

5B1. xCOUPLE, xCOUPID, xCORES, and xMARST

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 wave x, the deceased respondent is assigned a different xSUBHH, with xCOUPLE=1, while the surviving spouse or partner will be assigned xCOUPLE=5 (unless the surviving spouse or partner has remarried or re-partnered).

Labeled as "Wave Household ID," xCOUPID is created to link the sample members who are married or partnered in a given wave, or the sample members who were married or partnered in a prior wave but are no longer spouses or partners due to the death in the household. Specifically, 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 in the same way as if the couple had survived 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.

xCORES indicates whether a respondent lives with a spouse or partner. xMARST indicates whether a respondent is married, divorced/separated, widowed, or has never been married. In this version of the Tracker, these two variables are available only for HRS 2004 and forward. It is important to note that

xMARST may be different from the marital status assigned in the Core Final release. The marital status variable in the Tracker file (xMARST) was constructed by looking at several different sources of data, whereas 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 xCORES and 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.

For respondents in single-respondent households, we classify them as "living alone" (xCORES=3) if they were married (xMARST=1), and classify them as "living arrangements unknown" (xCORES=5) if they were divorced, separated, widowed, never married, or their marital status was unknown (xMARST>1.)

For respondents in two-respondent households, we classify them as "married spouses living together" (xCORES=1) if both respondents were married, and classify them as "living with someone regardless of marital status" (xCORES=2) if both respondents were not married.

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, whereas 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, respectively. 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) and EBB (2008) 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 HRS tracking information. It provides five 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). Persons for whom no report of death has been made are classified on the following principles. If the person 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 Tracker file and the public data use the last date of birth given, regardless of the source. xAge is only calculated for respondents who completed a core interview. This results in some discrepancies between xAGE and BIRTHYR.

5C3. *xNURSHM*

This variable indicates whether the respondent was in a nursing home 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, or by a spouse or partner describing the respondent's whereabouts.

There is no information available for an uncoupled respondent who was not interviewed by self or proxy, or for coupled respondents when neither of them gave an interview. The HRS cross-section sample weights (*xwgt*) are constructed to match the non-institutionalized population. Nursing home residents in a wave therefore receive a zero weight at that wave. By design, HRS does not interview baseline respondents who are nursing home residents, unless they are part of a couple whose other member is part of the non-institutionalized population. The exception to this was 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.

5C4. Changes to GNURSHM, HNURSHM, JNURSHM, KNURSHM and LNURSHM

In the process of finalizing the 2004 sample weights, we discovered some incorrect classifications in nursing home status for that wave. The problem was first detected when we noticed that 15 Early Baby Boomer (EBB) respondents were recorded as residing in nursing homes. A review of those cases indicated that 14 of the 15 had been misclassified (either due to data entry error or misclassification of an assisted living facility as a nursing home). The remaining EBB respondent was in a hospice rather than a nursing home. This led to a broader review of the 2004 cases, and ultimately of the 2008, 2006, 2004, 2002 and 2000 data.

Criteria used to identify 'suspicious' cases

We identified a set of key indicators that were highly predictive of misclassification in the 2004 wave. These include: 1) respondent type (self vs. proxy), 2) report of 'employee of institution' as a helper for ADLs and/or IADLs, and 3) prior wave nursing home status. We used this set of indicators to identify 'suspicious' cases in each of the 2000, 2002 and 2004 waves. Specifically, we flagged cases for review that were reported as residing in a nursing home in the current wave if the interview was conducted with a self-respondent, if the R did not report receiving help from an employee of the institution, and if the R was not recorded as residing in a nursing home in the prior wave. In addition, for the 2002 and 2004 waves, cases were flagged for review if there was disagreement between the core report and the interviewer observation report on nursing home status. No interviewer observations were available for HRS 2000.

Information used to reclassify cases

For all cases flagged for review, information was pulled from interview comments for nursing home-related questions in the core interview, as well as from interviewer call records. These comments were reviewed to make a determination as to whether the case should be reclassified. We also reviewed keystroke files for the flagged cases if the comments were inconclusive. We used the following decision rules for reclassifying cases from 'in nursing home' to 'not in nursing home':

1. The interviewer recorded a comment that indicated that the R was not residing in a nursing home or there was other evidence that the interviewer made a keying error (e.g., the interviewer tried to enter '0' in N115 and/or tried to back up and change the response in A023), or
2. The interviewer recorded a comment in an interview note or in the interviewer call records specifying that R lives in an assisted living

facility, or

3. The call record included comments indicating that the R lives in a single family home, or

4. The call record included comments indicating that the R was currently working.

The following decision rules were used to reclassify cases from 'not in nursing home' to 'in nursing home':

1. R was recorded as residing in a nursing home in the Interviewer Observations, and

2. The interview comments or interviewer call record included comments indicating that R resides in a nursing home.

Respondents flagged as suspicious based on the above criteria for which neither the interview comments nor the interviewer call records contained evidence to contradict the core assignment were left with their original assignment.

Change in 2006 questionnaire

Starting with the 2006 wave of HRS, we added an interviewer check at the nursing home question in order to reduce the number of entry errors and misclassifications. If the interviewer entered '1' at A028, indicating that the R is currently residing in a nursing home, a text box appeared informing the interviewer that s/he reported that R is currently residing in a nursing home. If the entry was correct, the interviewer could suppress the signal and go on with the interview. If the entry was incorrect, the interviewer was instructed to change the response in A028. This interviewer check substantially reduced the number of misclassifications.

Reclassification outcomes

HRS 2010:

12 cases changed from in a Nursing Home to not in Nursing Home.

5 cases changed from not in Nursing Home to in a Nursing Home.

HRS 2008:

20 cases were changed from in a Nursing Home to not in Nursing Home.

1 case was changed from not in a Nursing Home to in a Nursing Home.

HRS 2006:

9 cases were changed from in a Nursing Home to not in Nursing Home.

1 case was changed from not in a Nursing Home to in a Nursing Home.

HRS 2004:

77 cases were changed from in a Nursing Home to not in Nursing Home.

4 cases were changed from not in a Nursing Home to in a Nursing Home.

HRS 2002:

97 cases were changed from in a Nursing Home to not in a Nursing Home.

7 cases were changed from not in a Nursing Home to in a Nursing Home.

HRS 2000:

23 cases were changed from in a Nursing Home to not in a Nursing Home.

It is important to note that the Core data were not changed. Therefore, there will be mismatches between the core waves of data and the Tracker file with regard to nursing home residency. The Tracker file should be considered definitive.

5D. HRS Sample Weights

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 2000-2010 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. Up through 2002, the household level and respondent level weights (xWGTHH and xWGTR) are post-stratified to the March Current Population Survey for the year of data collection. Starting in 2004, the much larger American Community Survey is being 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 or later) 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 xWHYOWGT 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 and later waves, xWHYOHWT explains why a household has a zero weight and xWHYORWT 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 this 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. Changes to the HRS Sample Weights

5D2A. 1995 Respondent Weight

Our review of the full series of sample weights revealed a problem with the 1995 respondent weight (for the AHEAD cohort). The weighted sample size for the 1995 AHEAD sample (born in 1923 or earlier) was larger than the published population estimates based on the Current Population Survey (see Table A1), and it was also out of line relative to the weighted sample sizes for the 1993 and 1998 waves. We would expect some decline in the weighted sample size in each subsequent wave due to mortality; however, as shown in Table A2, the 1995 weighted sample was very close in size to the 1993 weighted sample. In contrast, as shown in Table A3, the sum of the household weights showed the expected pattern of decline over the waves and the 1995 weighted HRS count matches the 1995 weighted CPS count very closely.

A detailed investigation of the 1995 sample weights revealed that the household weights were calculated correctly. However the respondent weights were post-stratified to the 1993 CPS, rather than to the 1995 CPS. As a result, the population counts used to adjust the HRS respondent sample were too high, resulting in erroneously large sample weights.

The 1995 respondent sample weights have now been adjusted using the 1995 CPS as the post-stratification source. The weighted counts and distributions for the old and new respondent weights are given in Table A4. The original 1995 respondent weights overstated the size of the community-dwelling population born 1923 and earlier by 16%, and overstated the mean age of that population by 0.07 years (78.90 vs. 78.83).

The original sample weights were too large for each of the four birth ranges as shown in the ratios in the far right column, however the differential is most pronounced for the oldest two birth cohorts, for which mortality is highest. Thus, the oldest-old (born in 1913 or earlier) will be slightly overrepresented relative to those born between 1914 and 1923 in analyses based on the original 1995 sample weights. The impact of this error in the original weights is likely to be minimal for multivariate analyses that adjust for age; however, it could be more significant for descriptive analyses, e.g., prevalence estimates for health conditions, that are not age-adjusted. Also, analyses that present estimates of population counts for the AHEAD cohort based on 1995 data (e.g., the number of persons age 72 or over living with diabetes) will be over-estimated.

In revising the 1995 weights, we made use of the most recent information available on birth year and other eligibility indicators for the HRS sample. This resulted in some changes in respondent and household eligibility. The most significant change is that 55 respondents who had non-zero values on the original respondent weights are not cohort eligible (they were born in 1924 or later) and are, thus, assigned zero values on the revised respondent

weights. Conversely, three respondents who were assigned zero weights originally are now determined to be eligible and have non-zero values on the revised weights. At the household level, five households changed from eligible on the original weights to non-eligible (deceased) on the revised weights, and one household changed from non-eligible to eligible.

5D2B. 2004 Household and Respondent Weights

As noted in the sample design overview, Black and Hispanic households were oversampled in the HRS, AHEAD, EBB, and MBB cohorts. These oversamples are achieved by oversampling high density minority segments and subsampling (i.e., selecting only a random subset of) non-minority households within those segments. The weights are designed to account for the oversampling of minority respondents, along with other factors that lead to differential selection probabilities.

Our review of the sample weights revealed that the original baseline sample weights for the EBB cohort had not been adjusted to account for the subsampling of non-minority respondents in high density minority areas—non-minority respondents were treated as if their selection probabilities had been the same as minority respondents in those areas, whereas in fact they had only half the probability of being included. The 2004 weights have now been corrected to account for this. As expected, the main impact occurred for non-minority respondents in the EBB cohort. For some of this subgroup (those living in areas of high minority densities) the original weights were too small. Because the sample weights are generated for the entire sample concurrently, the weights for other respondents also changed slightly. Correlations between the original and revised weights, as well as comparisons on several key parameters are presented in Tables B1-B4. Figures B1 and B2 plot the original and revised weights at the household and respondent-level, respectively.

The corrections resulted in minor overall changes to the weights. The correction resulted in a difference of less than 1 percent in the household weight for 63% of households and less than 5 percent for 83% of households. At the respondent level, the difference in the weights was less than 1 percent for 65% of respondents and less than 5 percent for 86% of respondents.

We used the latest, most accurate information available on birth date and coupleness status to generate the revised weights. As a result, eligibility changed for a small number of households and respondents. Eight people in six HHs that were determined to be eligible when the original 2004 weights were calculated are now ineligible. At the respondent level, ten respondents that were originally determined to be eligible are now known to be ineligible. The net result is six fewer eligible households and ten fewer eligible respondents for the revised versus original weights.

As shown in Table B1, the correlations between the original and revised weights are extremely high. The lowest correlation is found for the EBB cohort, particularly those in the White/other race/ethnicity group.

Tables B2 and B3 show the sample size, mean, median, variance, and coefficient of variation for the original and revised weights, separately for the household weights (Table B2) and the respondent weights (Table B3). These statistics are shown in total, by cohort and, for the EBB cohort, by race and ethnicity.

There is very little shift in the distribution of the weights for all cohorts except the EBB cohort. Even for the EBBs, the shift is fairly modest. The mean and median shift upwards slightly, and the variance is reduced. Again, the distributional shift is more pronounced for the White/other group in the EBB cohort. These patterns are further illustrated in Figures B1 and B2, which plot the original weight against the revised weight at the household and individual level separately for Black, Hispanic and non-minority EBBs and for non-EBBs. Black and Hispanic EBBs and non-EBBs tend to cluster around the diagonal, indicating that the original and revised weights are very close. It is only non-minority EBBs for which the cluster departs from the diagonal, and for most of those cases, the revised weight is larger than the original weight. Even for this group, however, the correlations between the original and revised household and respondent weights are very high, as shown in Table B1.

A major concern for users will be what impact these changes have on the results of substantive analyses that were based on the original weights. Tables B4-B5 provide some insight into this issue. Table B4 presents weighted distributions and standard errors for key demographic, health, and economic indicators for both the total 2004 sample and the EBB cohort, for which the correction resulted in the largest change in weights.

As shown here, the distributions and standard errors based on the original and revised weights are extremely close. The primary difference is found for the respondent-level race/ethnic distribution for the EBB cohort, for which the weighted percent is slightly higher for the White/other group and slightly lower for the two minority groups using the revised weight. This difference is not statistically significant. Where the standard errors differ between the original and revised weights, they tend to be slightly lower based on the revised weights.

Table B5 presents distributions and standard errors for the same set of indicators for two additional subgroups—first the total White/other sample and second for the EBB White/other sample. The latter group is the group for which we would expect to see the largest differences. For the total White/other sample in 2004, the distributions and means for all of the indicators are very close. The same is true for most indicators for the White/other sample in the EBB cohort. The percentages with health insurance coverage, home ownership and IRA/Keogh accounts and the estimates of mean income and net worth are slightly lower based on the revised weight compared to the original weight, but none of these differences are statistically significant. For other indicators, the distributions are essentially identical.

Additional comparisons of research findings using the original versus revised weights based on replications of published and unpublished studies are presented in Section F.

5D2C. 2006 and 2008 Household and Respondent Weights

Because the baseline household weight is used as the starting point for generating sample weights in each subsequent wave, the 2006 and 2008 weights were also affected by the error in the 2004 EBB weights. We have updated and replaced all of these weights on the tracker file.

The impact on the 2006 and 2008 weights was even smaller than that for the 2004 weights. Table C1 presents correlations between the original and revised weights. Correlations for the total sample are 0.995 or higher for both waves, and within subgroups, the lowest correlation is 0.946. Correlations between the original and revised weights are high for the EBB cohort in both waves.

As with the 2004 weights, eligibility changed for a small number of households and respondents in 2006 and 2008.

5D2D. 2006 ACS Revision and Impact on HRS Weights

As noted above, the Census Bureau determined that the way in which the age perturbation was handled for several years of the Current Population Survey (CPS) and American Community Survey (ACS) was incorrect and it led to a distortion in the sex ratios particularly at the older ages. The Census Bureau has since released a new version of the 2006 ACS, which corrects for the error. The Census Bureau determined that the error in the 2004 CPS data had minimal impact on the data, however, and they opted not to release a new version of the 2004 CPS. The new ACS data was used in the revision of the 2006 HRS weights. Analyses comparing the sample weights based on the old and new ACS data showed that the change had very little impact on the HRS weights. As shown in Table D1, the correlations between the weights using the old and new ACS data are extremely high for all age and sex groups.

5D2E. Revisions to Supplement Weights

Sample weights for the supplemental studies (mail and Internet surveys) and components of the core survey (physical measures, biomarkers, leave-behind questionnaires) are based on the core sample weights. As a result, any supplement weights that relied on core weights from 2004, 2006 and 2008 had to be revised. The revisions were made by first multiplying the original supplement weight by the ratio of the revised to original core sample weight. This initial adjustment scaled the supplement weight appropriately. The supplement weights were then post-stratified to the weighted core sample (based on the revised core weights) from the prior core wave. The post-stratification adjustment was based on age, gender and race/ethnicity.

The revisions to the supplement weights resulted in minor changes in the weights for most respondents. Table E1 provides distributional statistics for the original and revised supplement weights, along with correlations between the two weights.

5D2F. Replication of Published and Unpublished Analyses

To further evaluate the impact of revisions to the sample weights, we are attempting to replicate analyses that have been conducted by other researchers. Results from one replication of analyses presented in a paper by Zivin et al. (2010) are provided in Tables F1 and F2. The paper is based on respondents who participated in the 2005 Prescription Drug Study, a mail survey of a subsample of HRS respondents. The analysis focused on medication non-adherence in this sample (n=3,071). Tables F1 and F2 present odds-ratios and confidence intervals for regression analyses based on the original and revised 2005 PDS weights. There are marginal changes in significance around the $p < .05$ level for a few odds-ratios (in bold), but the revised weights do not result in any changes to the substantive findings of the study.

A second replication is based on unpublished analyses of total assets conducted by Gretchen Lay at the University of Michigan. Table F3 provides a

comparison of mean assets by asset percentile for the 2004, 2006 and 2008 waves based on the original and revised weights. The impact of the revised weights on total assets is small overall, but there is some variation across waves. The impact is larger for the 2004 asset distribution than for 2006 or 2008. In addition, for 2004 the revised weights lead to slightly lower percentile estimations, whereas in 2006 and 2008 they lead to slightly higher percentile estimations.

Results from other replications will be added to this report as they become available. We welcome any contributions from researchers who wish to replicate their own work.

5D2G. References

Heeringa, Steven G., *Technical Description of the Asset and Health Dynamics Among the Oldest Old (AHEAD) Study Sample Design*, [1995]
(<http://hrsonline.isr.umich.edu/sitedocs/userg/AHDSAMP.pdf>)

Heeringa, Steven G.; Connor, Judith, *Technical Description of the Health and Retirement Study Sample Design*, [1995]
(<http://hrsonline.isr.umich.edu/sitedocs/userg/HRSSAMP.pdf>)

HRS website. 2011a. Sampling Weights: Revised for Tracker 2.0 and Beyond.
<http://hrsonline.isr.umich.edu/sitedocs/wghtdoc.pdf> (accessed April 25, 2011).

HRS website. 2011b. HRS Sample Evolution: 1992-1998.
<http://hrsonline.isr.umich.edu/sitedocs/surveydesign.pdf> (accessed April 25, 2011).

Zivin, K.; Ratliff, S.; Heisler, M. M.; Langa, K. M.; Piette, J. D. 2010
"Factors influencing cost-related non-adherence to medication in older adults: A conceptually based approach," *Value in Health*, 13:4, p338-345 [2010]

Table A1. Weighted samples sizes for persons born in 1923 or earlier: 1995 HRS vs. 1995 CPS

Birth range	1995 HRS		1995 CPS		Ratio HRS/ CPS
	Weighted N	Percent	Weighted N	Percent	
<=1923	22,248,170	100.0	19,219,159	100.0	1.16
<1909	2,510,034	11.3	2,149,321	11.2	1.17
1909-13	4,687,751	21.1	3,377,709	17.6	1.39
1914-18	6,197,476	27.8	5,542,565	28.8	1.12
1919-23	8,852,909	39.8	8,149,564	42.4	1.09

Table A2. Weighted sample sizes for the AHEAD cohort, 1993-2000

Birth range	Weighted respondent counts			
	1993	1995	1998	2000
<=1923	22,264,875	22,248,170	15,250,770	12,674,325
<1909	2,866,302	2,510,034	1,001,384	559,657
1909-13	4,433,396	4,687,751	2,378,348	1,850,112
1914-18	6,348,496	6,197,476	4,670,702	3,847,808
1919-23	8,616,681	8,852,909	7,200,336	6,416,748

Table A3. Weighted household sample for the AHEAD cohort: 1993-2000

Birth range	Weighted household counts				
	1993	1995 HRS	1998	2000	1995 CPS
<=1923	17,534,877	15,588,700	11,929,586	9,926,973	15,635,192

Table A4. Weighted respondent sample sizes and distributions for the AHEAD cohort based on original vs. revised respondent sample weights.

Birth range	Original weights		Revised weights		Ratio revised/ original
	Weighted N	Percent	Weighted N	Percent	
<=1923	22,248,170	100.0	19,220,715	100.0	0.86
<1909	2,510,034	11.3	1,961,110	10.2	0.78
1909-13	4,687,751	21.1	3,572,708	18.6	0.76
1914-18	6,197,476	27.8	5,642,589	29.4	0.91
1919-23	8,852,909	39.8	8,044,308	41.8	0.91

Table B1. Correlation between original and revised 2004 sample weights

Weight Cohort	Total	Hispanic	Black	White/ other
Household weight				
Total	0.9770	0.9946	0.9943	0.9707
AHEAD	0.9999	--	--	--
CODA	0.9993	--	--	--
HRS	0.9987	--	--	--
War Baby	0.9985	--	--	--
EBB	0.8878	0.9841	0.9813	0.8238
Respondent weight				
Total	0.9798	0.9919	0.9939	0.9753
AHEAD	0.9999	--	--	--
CODA	0.9991	--	--	--
HRS	0.9988	--	--	--
War Baby	0.9991	--	--	--
EBB	0.8894	0.9622	0.9753	0.8376

Table B2. Sample sizes and distributional statistics for original and revised 2004 household weights

	N	Mean	Median	Variance	Coeff var (%)	Sum of weights	N	Mean	Median	Variance	Coeff var (%)	Sum of weights
	Original weights						Revised weights					
Total	13078	4406	3742	7517029	62.22	57624655	13072	4408	3751	7225680	60.98	57618170
Cohort												
AHEAD	2440	3466	3557	2621457	46.71	8458145	2439	3466	3549	2570806	46.26	8454337
CODA	1236	4263	3745	2218608	34.94	5269396	1236	4262	3732	2162158	34.5	5267335
HRS	5854	3030	3083	1922313	45.75	17739325	5852	3022	3074	1837189	44.85	17685014
War Baby	1393	8685	8107	6339178	28.99	12098273	1393	8667	8076	5978524	28.21	12072636
EBB	2155	6524	5771	9590624	47.47	14059516	2152	6570	6112	8221210	43.64	14138848
Race/ethnicity												
Hispanic	1353	3393	2483	6052941	72.51	4590776	1352	3385	2455	5982617	72.27	4576022
Black	2121	2993	2059	4662038	72.14	6348225	2118	2978	2099	4452711	70.86	6307637
Other	9604	4861	4012	7562370	56.57	46685654	9602	4867	4047	7204166	55.15	46734511
EBB sample												
Hispanic	357	4603	3848	4841110	47.8	1643155	356	4434	3639	4747154	49.14	1578341
Black	424	4711	4339	4538559	45.22	1997353	422	4448	3966	4207025	46.11	1877101
Other	1374	7583	7329	9295103	40.21	10419008	1374	7775	7469	6340882	32.39	10683406

Table B3. Sample sizes and distributional statistics for original and revised 2004 respondent weights

	N	Mean	Median	Variance	Coeff var (%)	Sum of weights	N	Mean	Median	Variance	Coeff var (%)	Sum of weights
	Original weights						Revised weights					
Total	18588	4395	3720	7542620	62.49	81691803	18578	4395	3739	7290354	61.43	81651345
Cohort												
AHEAD	3006	3533	3614	2742186	46.87	10620486	3005	3534	3608	2701630	46.51	10618693
CODA	1722	3963	3057	2084519	36.43	6824051	1722	3961	3048	2038052	36.04	6820770
HRS	9012	3067	3169	18885430	44.77	27638321	9010	3060	3160	1812419	43.99	27572918
War Baby	2158	8647	8188	6345286	29.13	18659331	2156	8641	8162	6139088	28.67	18629891
EBB	2690	6673	5932	9771235	46.85	17949614	2685	6707	6239	8353964	43.09	18009073
Race/ethnicity												
Hispanic	1724	3274	2328	5678564	72.77	5645161	1720	3256	2372	5591495	72.63	5599941
Black	2668	2941	2009	4418588	71.47	7846984	2664	2928	2048	4233431	70.28	7799452
Other	14193	4804	3966	7638467	57.53	68199658	14191	4809	3988	7337058	56.33	68251952
EBB sample												
Hispanic	394	4571	3912	4465792	46.24	1800783	391	4366	3577	4350000	47.77	1707101
Black	475	4717	4442	3895929	41.85	2240344	473	4442	4026	3615657	42.81	2101099
Other	1821	7638	7530	9570377	40.50	13908487	1821	7798	7431	6747597	33.31	14200873

Figure B1. Scatterplot of original vs. revised HRS 2004 household weights

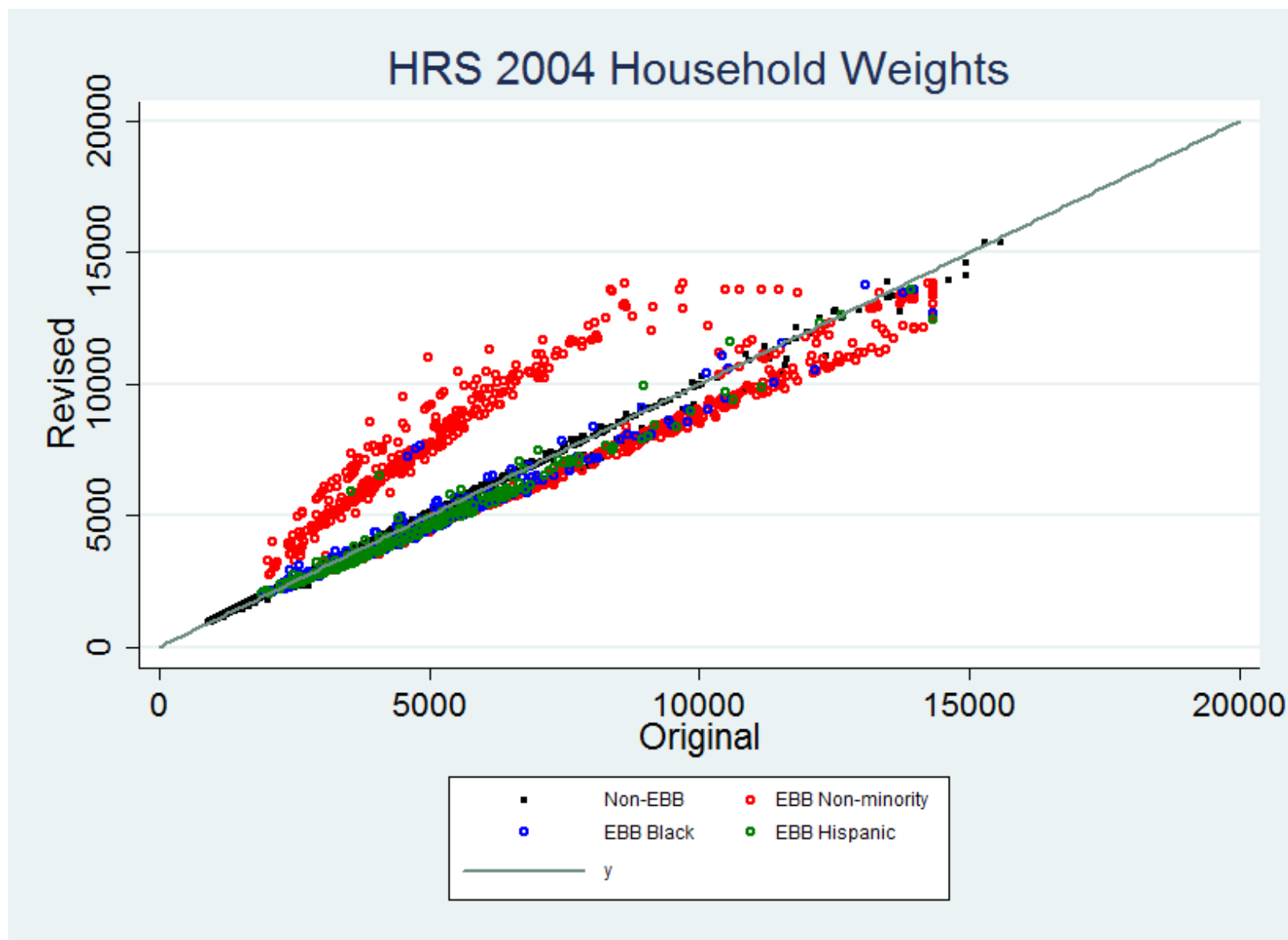


Figure B2. Scatterplot of original vs. revised HRS 2004 respondent weights

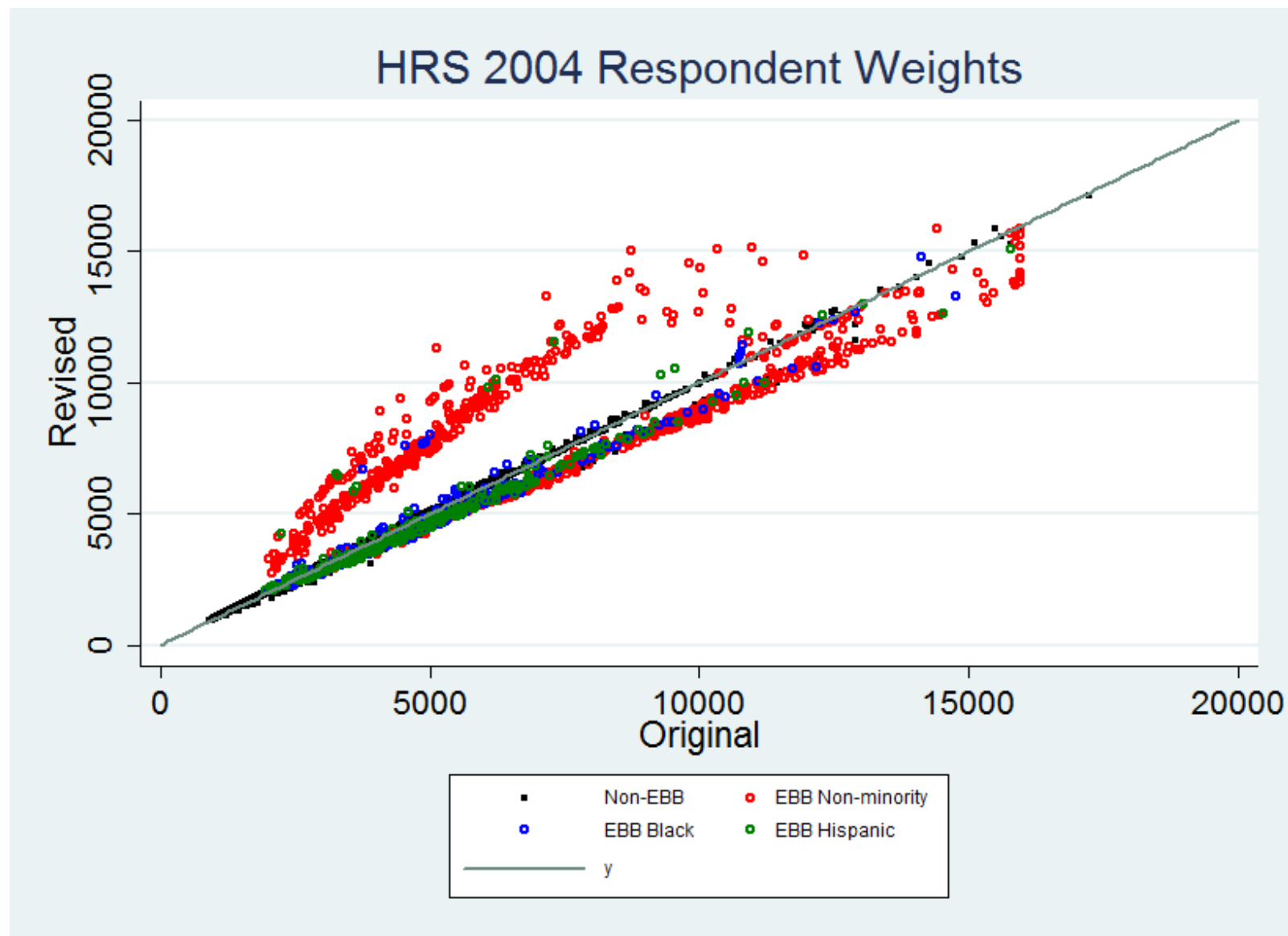


Table B4. Weighted estimates for substantive variables using old vs. new weights for total sample and EBB cohort

	Total sample				EBB cohort			
	Original wt		Revised wt		Original wt		Revised wt	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
<i>Respondent indicators</i>								
Age group								
< 55	20.88	0.59	20.81	0.59	78.57	1.10	78.33	1.09
55-59	20.17	0.46	20.22	0.46	21.43	1.10	21.67	1.09
60-64	15.59	0.36	15.59	0.36	--	--	--	--
65-69	12.19	0.25	12.19	0.25	--	--	--	--
70-74	10.66	0.25	10.66	0.25	--	--	--	--
75+	20.51	0.62	20.52	0.62	--	--	--	--
Gender								
Male	45.92	0.31	45.93	0.31	55.39	0.88	55.29	0.86
Female	54.08	0.31	54.07	0.31	44.61	0.88	44.71	0.86
Race/ethnicity								
Hispanic	6.90	0.82	6.84	0.81	9.94	2.02	9.39	1.84
Black	9.55	0.51	9.50	0.50	12.39	1.10	11.58	1.00
White/other	83.55	0.90	83.66	0.88	77.67	2.16	79.03	1.95
Education								
< 12 years	20.12	0.73	20.12	0.72	11.36	1.39	11.09	1.29
12 years	32.89	0.56	32.89	0.56	26.08	1.13	26.24	1.19
13-15 years	22.31	0.45	22.31	0.46	28.69	1.13	28.71	1.14
16+ years	24.68	0.79	24.68	0.78	33.86	1.73	33.96	1.75
% working for pay	46.77	0.63	46.73	0.62	78.13	1.13	78.02	1.03
% unemployed	1.38	0.13	1.38	0.12	3.95	0.49	3.90	0.46
% with hypertension	50.07	0.53	50.13	0.53	34.77	1.13	34.92	1.08
% with diabetes	16.37	0.33	16.45	0.33	11.39	0.63	11.64	0.68
% with heart disease	72.02	0.83	71.66	0.41	10.61	0.61	10.58	0.61
% with health insurance	72.02	0.83	71.99	0.80	76.21	1.75	76.24	1.58
<i>Household indicators</i>								
% who own primary residence	77.43	0.60	77.40	0.59	76.75	1.30	76.76	1.28
% who have IRA/Keogh	39.28	0.79	39.10	0.78	40.22	1.76	39.75	1.73
Mean income	61,777	1,480	61,658	1,435	84,667	3,293	84,202	3,151
Mean net worth	427,145	19,422	423,379	19,210	386,357	40,793	373,560	37,289

Table B5. Weighted estimates for substantive variables using old vs. new weights for selected subsamples

	Total White/other				EBB White/other			
	Original wt		Revised wt		Original wt		Revised wt	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
<i>Respondent indicators</i>								
Age group								
< 55	19.60	0.73	19.80	0.71	78.63	1.33	78.37	1.28
55-59	19.74	0.52	19.78	0.52	21.37	1.33	21.63	1.28
60-64	15.53	0.42	15.46	0.42	--	--	--	--
65-69	12.37	0.28	12.32	0.28	--	--	--	--

70-74	11.05	0.30	11.00	0.29	--	--	--	--
75+	21.71	0.74	21.62	0.73	--	--	--	--
Gender								
Male	46.55	0.34	46.54	0.34	56.93	1.02	56.61	0.99
Female	53.45	0.34	53.46	0.34	43.07	1.02	43.39	0.99
Education								
< 12 years	15.45	0.46	15.44	0.46	5.81	0.58	5.95	0.57
12 years	34.42	0.67	34.41	0.68	26.74	1.39	26.92	1.46
13-15 years	23.00	0.48	23.02	0.49	29.08	1.30	29.02	1.31
16+ years	27.13	0.88	27.13	0.87	38.37	1.97	38.11	1.97
% working for pay	47.03	0.74	47.09	0.73	80.79	1.27	80.38	1.17
% unemployed	1.17	0.14	1.18	0.14	3.26	0.53	3.26	0.50
% with hypertension	48.50	0.64	48.54	0.64	31.55	1.45	32.06	1.35
% with diabetes	14.66	0.35	14.73	0.35	9.76	0.72	10.18	0.75
% with heart disease	22.41	0.47	22.36	0.46	10.60	0.72	10.60	0.72
% with health insurance	76.67	0.74	76.57	0.73	82.62	1.47	81.97	1.33
<i>Household indicators</i>								
% who own primary residence	81.23	0.58	81.08	0.56	83.35	1.36	82.62	1.28
% who have IRA/Keogh	44.58	0.78	44.31	0.77	48.01	1.93	46.69	1.90
Mean income	67,117	1,655	66,924	1,632	96,113	3,642	94,292	3,694
Mean net worth	494,104	22,390	488,719	22,216	471,290	51,758	446,658	46,389

Table C1. Correlation between original and revised 2006 and 2008 sample weights

Weight Cohort	2006	2008
Household weight		
Total sample	0.9958	0.9967
Cohort		
AHEAD	0.9974	0.9983
CODA	0.9753	0.9878
HRS	0.9952	0.9974
War Baby	0.9679	0.9703
EBB	0.9996	0.9999
EBB cohort		
Hispanic	0.9955	0.9999
Black	0.9996	0.9996
White/other	0.9999	1.0000
Respondent weight		
Total sample	0.9910	0.9949
Cohort		
AHEAD	0.9967	0.9982
CODA	0.9725	0.9903
HRS	0.9913	0.9960
War Baby	0.9461	0.9598
EBB	0.9923	0.9993
EBB cohort		
Hispanic	0.9882	0.9996
Black	0.9907	0.9995
White/other	0.9901	0.9990

Table D1. Correlation between original and updated 2006 HRS sample weights (update based on new 2006 ACS data)

Age group	Household	Respondent	
		Male	Female
< 55	0.9999	0.9998	0.9999
55-59	0.9999	0.9999	0.9999
60-64	0.9999	0.9998	0.9999
65-69	0.9996	0.9995	0.9995
70-74	0.9990	0.9990	0.9990
75-79	0.9965	0.9975	0.9965
80-84	0.9982	0.9986	0.9982
85+	0.9989	0.9995	0.9990
Total	0.9998	0.9968	0.9970

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. The sample weights for each

component are provided in the 2012 Tracker file, along with sample selection indicators. 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:

PSYSOC04 - sample selection indicator for 2004 psychosocial questionnaire
JWGTR_PS - final respondent weight for 2004 psychosocial subsample

Disability Leave-Behind:

DISAB04 - sample selection indicator for 2004 disability questionnaire
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:

http://hrsonline.isr.umich.edu/modules/meta/tracker/desc/PMWeight2004_Description_public.pdf

Leave-Behind Questionnaires:

http://hrsonline.isr.umich.edu/modules/meta/tracker/desc/LBWeights2004_Description_public.pdf

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 5th and 95th 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 2012 Tracker file along with sample selection and eligibility indicators. These include:

EFTFASSIGN - enhanced face-to-face sample indicator
KPMWGTR - final respondent weight for 2006 physical measures subsample
KBIOWGR - 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:
<http://hrsonline.isr.umich.edu/sitedocs/userg/dr-011.pdf>

Information on the biomarker collection can be found in the Data Description for the 2006 Biomarker Data file:
<http://hrsonline.isr.umich.edu/modules/meta/bio2008/desc/biomkr06dd.pdf>

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 5th and 95th 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
PSYSOC06 - sample selection indicator for 2006 psychosocial questionnaire
KLBWGTR - final respondent weight for the 2006 leave behind questionnaire subsample

More information about the questionnaire can be found in the following User Guide:
<http://hrsonline.isr.umich.edu/sitedocs/userg/HRS2006LBQscale.pdf>

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

Sample weights for the 2008 Physical Measures and Psychosocial Leave Behind Questionnaire were developed using the same methodology as in 2006. Sample weights for the 2008 biomarkers are forthcoming. The sample weights for each component are provided in the 2008 Tracker file along with sample selection and eligibility indicators. These include:

EFTFASSIGN - enhanced face-to-face sample indicator
PSYSOC08 - sample selection indicator for 2008 psychosocial questionnaire
LPMWGTR - final respondent weight for the 2008 physical measures subsample
LLBWGTR - final respondent weight for the 2008 leave behind questionnaire subsample

6. Important Data Changes in Early Version 1.0 of Tracker 2012

6A. Race, Hispanic, and Gender Updated

In the last year we have done a review of some of the basic demographic variables. In that process we discovered that RACE, HISPANIC and GENDER were incorrect for some cases in earlier versions of Cross Wave Tracker file. As a result, 637 respondents have changes to Race, 454 respondents have changes to Hispanic, and 5 respondents have changes to gender.

7. Obtaining the Data

7A. Registration and Downloading the Data

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

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

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

- o not attempt to identify respondents

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

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

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, Room 3050
P.O. Box 1248
Ann Arbor, MI (USA) 48106-1248

Alternately, you may contact us by e-mail at 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 2012, Early Version 1.0 data. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

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, point your Web browser to the HRS Web site:

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

8B. Contact Information

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

Internet: Help Desk at our Web site

E-mail: hrsquestions@umich.edu

Postal service:

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

FAX: (734) 647-1186

APPENDIX

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, those with multiple identities (i.e., multiple "hhid" and "pn" combinations) across waves may 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 difficult because of the HRS-AHEAD overlaps explained earlier. 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 dataset HEALTH

9A1. Retaining the Entire, Original HRS Sample

(i) SAS Code

```
data tracker;
  set dat.trk2010;
  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 trk2010
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\trk2010.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.trk2010 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 trk2010
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\trk2010.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.

(i) SAS Code

```

data tracker;
  set dat.trk2010;
  if ovresult=8 and hhid not eq "205864" then hhid=ovhhid;
  if ovresult=8 and hhid not 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 trk2010
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\trk2010.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'.

```