RAND Center for the Study of Aging

A RAND LABOR AND POPULATION PROGRAM

RAND HRS Fat Files

The RAND HRS Fat Files contain most of the raw HRS/AHEAD variables merged to the Respondent-level. There is one Fat File per year for 1992, 1993, 1994, 1995, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, and 2016, and each Fat File is sorted by HHIDPN. The files correspond to those used in developing RAND HRS Longitudinal File.

The RAND HRS Fat Files contain all the raw variables collected at the Respondent or household-level, except those from "Other Person" modules (i.e., data on children, siblings, household members, helpers, and transfers between Respondents and their children). Users should also note the following:

- The 1992 and 1994 Fat Files include all the raw variables.

- The 2002 Fat File does not include the raw variables from the pension preloads (PR_P) or the asset reconciliation (U_A) modules.

- The 2012-2016 Fat Files do not include the raw variables from the pension (PR_P) or jobs (PR_JB) preloads modules. The raw variables from the pension (J_P) module, however, are included in the Fat Files, and have been transposed to the Respondent level.

More detailed documentation is available on the RAND Center for the Study of Aging website:
http://www.rand.org/labor/aging/dataprod.

1. What makes these files different from the Public Release HRS Data?
2. Additions and Changes to the Public Release Data
3. Lining up the years
4. Overlap cases and merging with RAND HRS Data
5. Public Release Versions
6. Downloading the Files

1. What makes these files different from the Public Release HRS Data?

The Health and Retirement Study (HRS) surveys collect a considerable amount of information about elderly households and use a complicated questionnaire to minimize the burden on any single Respondent within the household. For example, the Respondent who answers the household finance module may not be the same person who answers the module on children and other household members. This format complicates the way the data are structured upon its public release. For instance, since only one person in a couple household provides financial and family information, the raw HRS data provides these data at the household-level, but many analyses use Respondent-level observations.

We have created public use files which simplify this data structure in several ways. We have reorganized the data so that each observation represents one Respondent, and we have merged the appropriate information from the various modules together. Thus, household-level information is present for each individual Respondent.

Similarly, certain Respondent-level information is collected by asking one household member to answer about another. For example, the Financial Respondent provides earnings information for both her/himself, and her/his spouse. We have reassigned this data to "self and "spouse/partner" variables for both Respondents to clarify to whom the information applies.

Finally, we make the files easy to merge across survey waves and with the RAND HRS Longitudinal File, and apply consistent variable naming conventions across waves that reflect those already used by the HRS data.

2. Additions and Changes to the Public Release HRS Data

As previously noted, the RAND HRS Fat Files include most of the public release data. "Other Person" modules containing data on children, siblings, household members, helpers, and transfers between Respondents and their children are not included in the RAND HRS Fat Files.

However, the RAND Center for the Study of Aging also produces the RAND HRS Family Data that does include the "Other Person" modules. The RAND HRS Family Data contains a cleaned, processed, and streamlined collection of variables related to the Respondent's family. The files described here include a subset of available characteristics of all children of HRS Respondents and spouses.

Variables are added:

- Identifying variables
- Clarify the household level data as it applies to Respondent level data (for financial and family variables)
- Generally helpful variables or flag discrepancies
- Provide categorical bracket variables, which if not provided in the raw data, summarize unfolding bracket questions
- Correct raw data to reflect Data Alerts

Identifying variables: not wave-specific

Description	Identifying Variable Name in *RAND HRS Fat File*	Notes
HRS Household ID: Numeric	HHID	ID Fields are provided in both numeric and character format. HRS provides IDs in character format, but numeric formats are included in the *RAND HRS Fat File* to facilitate merging with data from previous waves.
HRS Household ID: Character	HHIDC	
HRS Household ID + Person Number: Numeric	HHIDPN	
HRS Household ID + Person Number: Character	HHIDPNC	
Person Number: Numeric	PN	
Person Number: Character	PNC	

Cohort	HACOHORT (derived)	HACOHORT is set to xZ023, exactly as is: • 1 = AHEAD • 2 = CODA (Children of depression) • 3 = HRS • 4 = War Babies (WB) • 5 = Early Baby Boomers (EBB) • 6 = Mid Baby Boomers (MBB) • 7 = Late Baby Boomers (LBB) [NOTE: HACOHORT indicates the sample in which R entered the study, e.g., HRS in 1992, AHEAD in 1993, CODA or WB in 1998 and EBB in 2004]
Cohort eligibility	RACOHBYR (derived)	RACOHBYR is based on BIRTHYR. It may differ from COHORT. • 0 = not eligible for any cohort (born after 1953) • 1 = AHEAD (born before 1924) • 2 = CODA (born 1924-1930) • 3 = HRS (born 1931-1941) • 4 = War Babies (born 1942-1947) • 5 = Early Baby Boomers (born 1948-1953) • 6 = Mid Baby Boomers (born 1954-1959) • 7 = Late Baby Boomers (born 1960-1966)

Identifying variables: wave-specific

Description	Identifying Variable Name in *RAND HRS 2010 Fat File*	Notes
2010 Sub-Household Identifier	MSUBHH (char)	ID Fields are provided in both numeric and character format. HRS provides IDs in character format, but numeric formats are included in *the RAND HRS Fat files* to facilitate merging data across years.
	MSUBHHN (numeric)	
2010 Household Identifier	MHHIDN (numeric) = (HHIDx10) +	

	MSUBHHN	
Family Respondent	MFAMR	MFAMR is the categorical variable from the **core** data: • 1 = R is FamR, only one FamR in HHold • 3 = R is second FamR, answers not retained • 5 = R is not FamR, at most one FamR in HHold
Financial Respondent	MFINR	MFINR is the categorical variable from the **core** data: • 1 = R is FinR, only one FinR in HHold • 3 = R is second FinR, answers not retained • 5 = R is not FinR, at most one FinR in HHold
Respondent in nursing home	MNHM	JNHM is derived from core data.

Assignment of Income Variables to Self/Spouse-Partner Variables (FR/FP)

In a couple household, the Financial Respondent (FR) provides information about the household's finances. For example, the Financial Respondent provides information regarding earned income for both her/himself, as well as for her/his spouse (FP). In the raw HRS modules, the Respondent earned income data apply to the Financial Respondent, and the spouse/partner earned income data apply to the Financial Respondent's spouse. In the RAND HRS Fat Files, we assign these data, collected at the household level, to Respondent-level "self" and "spouse/partner" variables for both Respondents to clarify to whom the information applies.

In the RAND HRS 2010 Fat Files, variables that begin with MQR apply to the Respondent ("self"), whereas variables that begin with MQP apply to the Respondent's spouse ("spouse/partner"). Note that in 2010, MQ are the first two letters for the financial variables. Generally, if an individual is the Financial Respondent, the MQnnn variables for the Financial Respondent are assigned to MQRnnn, and the MQppp variables for the Financial Respondent's spouse are assigned to MQPnnn (not MQPppp). We always keep the Respondent variable stem so that it is easier for the user to pick the correct spouse variable. If a Respondent is the spouse/partner of the Financial Respondent, then the MQnnn variables for the Financial Respondent are assigned to MQPnnn, and the MQppp variables for the Financial Respondent's spouse are assigned to MQRnnn. The table below summarizes these assignments.

Variable Name Assignments			
RAND HRS 2010 Fat File			
Original Variable		**Reassigned Variable (FR)**	
FinR	**Finr's Spouse**	**Self**	**Spouse/Partner**
MQnnn	MQppp	MQRnnn	MQPnnn
MQ020	MQ045	MQR020	MQP020

For example, the Financial Respondent's report of earnings from wages and salaries for her/himself is in MQ020, and for her/his spouse in MQ045. In the observation for the Financial Respondent, MQR020 = MQ020 and MQP020 = MQ045. In the observation for the Financial Respondent's spouse, MQR020 = MQ045 and MQP020 = MQ020.

Income Variables Assigned to Self/Spouse-Partner Variables (FR/FP)
Recoded, Reassigned, or Combined versions of variables

1. Who worked for pay last year?

The questions on who worked for pay last year are different depending on whether the Respondent (and spouse) are over 65 years old:

- MQ010: Did you (or your spouse) do any work for pay last year?
- MQ011: Which one of you worked for pay last year?
- MQ012: Did you do any work for pay last year?
- MQ013: Did your spouse do any work for pay last year?

We make 4 new variables that are not dependent on the Financial Respondent, or the age of the Financial Respondent's spouse:

Variable *description*	If individual is FinR	If individual is NOT FinR
MQR012 *R did any work for pay last year*	**l = Yes** if single R and MQ010=l, MQ011=l or 3 (FinR or both), or MQ012=1 **5 = No** if MQ010=5, or MQ011=2 (spouse), or MQ012=5 **8 = DK** if MQ010=8 or MQ012=8 **9 = RF** if MQ010=9 or MQ012=9	**l = Yes** if MQ011=2 or 3 (spouse or both), or MQ013=l **5 = No** if MQ010=5, or MQ011=l (FinR), or MQ013=5 **8 = DK** if MQ010=8 or MQ013=8 **9 = RF** if MQ010=9 or MQ013=9
MQP012 *Spouse/Partner did any work for pay last year*	**l = Yes** if MQ011=2 or 3 (spouse or both), or MQ013=l **5 = No** if MQ010=5, or MQ011=1 (FinR), or MQ013=5 **8 = DK** if MQ010=8 or MQ013=8 **9 = RF** if MQ010=9 or MQ013=9 **.Z** =inapplicable (no FinR spouse)	**l = Yes** if MQ011=l or 3 (FinR or both), or MQ012=l **5 = No** if MQ010=5, or MQ011=2 (spouse), or MQ012=5 **8 = DK** if MQ010=8 or MQ012=8 **9 = RF** if MQ010=9 or MQ012=9

Variable *description*	If individual is FinR	If individual is NOT FinR
MQR010 *Either R or spouse did any work for pay last year*	• = **1,** if MQR012=l or MQP012=l • = **5**, if MQR012=5 and, if FinR married/partnered, MQP012=5 • = **8** (DK), if MQR012=8 or MQP012=8 • = **9** (RF), if MQR012=9 or MQP012=9	
MQR011 *Who did any work for pay last year*	• = **l**, if MQR012=l and (single R or MQP012 not equal 1) • = **2**, if MQP012=1 and MQR012 not equal 1 • = **3**, if MQR012=1 and MQP012=l • = **8** (DK), if MQR012=8 or MQP012=8 • = **9** (RF), if MQR012=9 or MQP012=9	

2. Who receives transfer, IRA/pension, trust income?

Questions for each type of transfer income as well as retirement pension and annuity income begin as follows:

A. Was any income of this type received?
B. If yes, who received it?

In questions about IRA or Keogh accounts (the first and second largest) the question is asked, "Is that yours or your spouse's account?" For assets put into a trust, up to 4 relations of those receiving income from the trust can be specified, including the Financial Respondent and spouse.

For these questions where the Financial Respondent or spouse can be specified in the answer, we create an FRnnnn version, which is recoded to be relative to "self." For example, MQ065 indicates whether the Financial Respondent, Financial Respondent's spouse, or both received Unemployment Compensation last year:

• 1 = FinR received
• 2 = Spouse/Partner received
• 3 = Both received
• 8, 9 = DK, RF

For the Financial Respondent, MQR065 is set to MQ065 exactly as it is. For the Financial Respondent's spouse, MQR065 is set using the following recoding: MQR065 = 2, if MQ065=l; MQR065= 1, if MQ065=2.

The variables listed below are recoded in this way when assigned to their Financial Respondent versions:

Original Variable	Self (FR) Variable	Description
MQ065	MQR065	WHO RECEIVED INC FROM UNEMPLOYMENT LCY
MQ075	MQR075	WHO RECEIVED FROM WORKERS COMP LCY
MQ106	MQR106	WHO RECEIVED INC FROM SSI LAST MONTH
MQ114	MQR114	WHO RECEIVED INC FR WELFARE NOT SSI LCY
MQ120	MQR120	WHO RECEIVES VET BEN MILITARY PENSION
MQ165_1	MQR165_1	WHO HAS IRA ACCOUNTS -1
MQ165_2	MQR165_2	WHO HAS IRA ACCOUNTS -2
MQ165_3	MQR165_3	WHO HAS IRA ACCOUNTS -3
MQ183	MQR183	WHO HAS OTHER IRA ANNUITIES
MQ216	MQR216	WHO REC PENSION RETIREMENT INC
MQ274	MQR274	WHO RECEIVED INC FROM ANNUITIES
MQ465M1 thru MQ465M5	MQR465M1 thru MQR465M5	RELATION TRUST BENEFITS REC

3. Whose parents an inheritance is from

If the Respondent or her/his spouse have received any inheritance, the Financial Respondent is asked:

"From whom was that inheritance received, that is what was their relationship to you?"

The relationships include the Financial Respondent's parents and spouse's parents. The questionnaire asks about up to 3 inheritances in variables MQ490_1- MQ490_3. We have created a MQRnnnn version, which is recoded to be relative to "self."

For example, for the Financial Respondent, MQR490_1=1 if the inheritance is from the Financial Respondent's parents (MQ490_1=14), or MQR490_1 =2 if it is from the Financial Respondent's spouse's parents (MQ490_1=34).

For the Financial Respondent's spouse, MQR490_1 is set to MQ490_1 with the following recoding:
- Set MQR490_1 to 2 (spouse's parents), if MQ490_1=l4 (Financial Respondent's parents)
- Set MQR490_1 to 1 (own parents), if MQ490_1=34 (Financial Respondent's spouse's parents)

The variables recoded in this way are:

Original	Self(FR)	Description
MQ490_1 - MQ490_3	MQR490_1 - MQR490_3	LUMP SUM - WHO FROM

3. Lining up the years

The table below summarizes the RAND HRS Fat Files, the first character of variable names, and RAND HRS survey waves associated with each interview year. References to "raw" data mean the data modules as publicly released by HRS.

Note that for years affected by overlap cases, there are two versions of the RAND HRS Fat Files, one with the original Respondent IDs from the raw public release data, and the other with the most recent IDs as found on the Tracker and RAND HRS Longitudinal File. For most cases, the IDs are identical on both files, and if you drop all overlap cases they are essentially the same.

In the raw public release data for 1992 and 1993, most variable names begin with "V". To make unique variable names when merging 1992 and 1993 data together, the variables in the 1993 Fat File that have names like Vnnnn, e.g., V1234, have been renamed to begin with "B", e.g., B1234.

Fat File names follow the following naming convention: [A/H][D]yy[E/F][vers]

- "A" is for Ahead and "H" is for HRS
- "D" if present, indicates the version of IDs that will merge best with other RAND HRS data products
- "yy" is the interview year
- "E" is for early release and "F" is for final release
- "vers" indicates the version of the HRS public release and of the RAND release

For example, hd10f5e is the RAND HRS Fat File for HRS 2010, HRS Final Release Version 5.1, RAND version E.

Interview Year	RAND HRS Fat File		Associated First Letter(s) of Variables (raw and Tracker)	RAND HRS Longitudinal File Wave
	IDs as in public release	*IDs as on Tracker, RAND HRS*		
1992	h92flb	hd92flb	**A** (tracker) and V (raw, fat files)	Wave 1
1993	a93f2a	ad93f2a	B (tracker, fat files) and V (raw)	Wave 2 (2A in codebook)
1994	h94fla		C (tracker) and W (raw, fat files)	Wave 2 (2H in codebook)
1995	a95f2b	ad95f2b	**D**	Wave 3 (3A in codebook)
1996	h96f4a		**E**	Wave 3 (3H in codebook)
1998	h98f2c	hd98f2c	**F**	Wave 4
2000	h00flc		**G**	Wave 5
2002	h02f2d		**H**	Wave 6
2004	h04flb		**J**	Wave 7
2006	h06f3a		**K**	Wave 8
2008	h08f3a		**L**	Wave 9
2010	h10f5e	hd10f5e	**M**	Wave 10
2012	h12f2a		**N**	Wave 11
2014	h14f2b		**O**	Wave 12
2016	h16e2a		**P**	Wave 13

4. Overlap cases and merging with RAND HRS Data

There are two versions of the 1992, 1993, 1995, 1998 and 2010 files, to accommodate the HRS-AHEAD (1992), AHEAD-AHEAD (1993, 1995, and 1998) and HRS-HRS (2010) overlap cases.

For all these years, one version has the original HHIDPN, as found in the HRS public release data, and the other has HHIDPN as assigned on the RAND HRS Data:

1992 Fat File variations
     h92flb: with original HRS IDs
     hd92flb: with IDs that merge best with RAND HRS

1993 Fat File variations
     a93f2a: with original HRS IDs
     ad93f2a: with IDs that merge best with RAND HRS

1995 Fat File variations
     a95f2b: with original HRS IDs
     ad95f2b: with IDs that merge best with RAND HRS

1998 Fat File variations
     h98f2c: with original HRS IDs
     hd98f2c: with IDs that merge best with RAND HRS

2010 Fat File variations
     h10f5e: with original HRS IDs
     hd10f5e: with IDs that merge best with RAND HRS

The "hd" and "ad" versions of the 1992, 1993, 1995, 1998 and 2010 data are available on the HRS website as they will easily merge with the RAND HRS data. Contact us at RANDHRSHelp@rand.org to request the versions with the original IDs.

Please see the RAND HRS Longitudinal File codebook, specifically Section A (Overlap Identifier for cases that moved from HRS to AHEAD), for background information on overlap cases. HRS/AHEAD overlap cases are identified by their AHEAD IDs in the RAND HRS Longitudinal File, and in the Tracker files. AHEAD-AHEAD overlap cases are identified by the most recent AHEAD ID in the RAND HRS Longitudinal File and in the Tracker file.

5. Public Release Versions

- HRS 1992: V1.01
- AHEAD 1993: V2.1
- HRS 1994: V1.0
- AHEAD 1995: V2.0
- HRS 1996: V4.0
- HRS 1998: V2.3
- HRS 2000: V1 .0 (with January and November 2005 HRS Data Alert applied)
- HRS 2002: V2.0 (dropped 2 cases based on Tracker Documentation)
- HRS 2004: V1.0
- HRS 2006: V3.0
- HRS 2008: V3.0
- HRS 2010: V5.1
- HRS 2012: V2.0
- HRS 2014: V2.0
- HRS 2016: E2.0

- Tracker 2.0 was used in developing the 1993 and 1994 fat files.
- Tracker 3.1 was used in developing the 1995, 1996, 1998, and 2000 fat files.
- Tracker 2004.vl was used in developing the 2002 and 2004 fat files.
- Tracker 2006.vl (released October 2007) was used in developing the 2006 fat files.
- Tracker 2008.vl (released October 2010) was used in developing the 2008 fat files.
- Tracker 2010.v1 (released April 2013) was used in developing the 2010 fat files.
- Tracker 2012.vl (released March 2015) was used in developing the 2012 fat files.
- Tracker 2014.v1 (released July 2017) was used in developing the 2014 fat files.
- Tracker Early 2016.v3 (released May 2019) was used in developing the 2016 fat files.

6. Downloading the Files

To access the data:

- All HRS and RAND HRS data products are available at: https://hrs.isr.umich.edu/data-products (please click this link)
- Click the "**Register and Access Public Data**" link in the bottom left-hand corner of the page.
- If you are not registered with HRS, please click the "**New Users**" link. Registration is free. You will receive a password within 24 hours.
- Once your account is active, please click the "**Registered Users**" link, and enter your **UserID** and **Password**.

- Click on the "**Data Downloads**" link, and you will be brought to a page that contains the various data sources that are available for download.
- Under the section titled, "HRS Biennial Data Products", you will see several links. The RAND HRS Fat Files contain "RAND HRS Fat File" in the title.
- Clicking on one of these links will bring you to a page with several ZIP files. Each one contains data that are already formatted for a particular statistical package (i.e., SAS, Stata, and SPSS).
- Files are available for 1992, 1993, 1994, 1995, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, and 2016, and each can be used with the current release of the RAND HRS Longitudinal file.

Please contact RAND HRS Help (RANDHRShelp@rand.org) with any questions.