

The RAND HRS Data (Version P)

August 2016

Data Distribution Description

1. Overview

1.1. Data Description

The RAND HRS Data file is a cleaned, easy-to-use, and streamlined version of the Health and Retirement Study (HRS) with derived variables covering a broad though not complete range of measures and named consistently across waves. The file includes imputations for income, assets, and medical expenditures developed at RAND. The development and continued maintenance of the RAND HRS Data is supported by the National Institute on Aging (NIA) and the Social Security Administration (SSA).

The HRS is a national panel survey of individuals aged 51 and above and their spouses. Its main goal is to provide panel data that enable research and analysis in support of policies on retirement, health insurance, saving, and economic well-being. The survey elicits information about demographics, income, assets, health, cognition, family structure and connections, health care utilization and costs, housing, job status and history, expectations, and insurance. HRS is mainly funded by NIA.

As of 2013, twelve HRS waves are available for study. The RAND HRS Data file is based on 1992, 1993, 1994, 1995, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012 final release and 2014 early release. The complete HRS includes six cohorts:

- HRS cohort, born 1931 to 1941, baseline 1992
- AHEAD cohort, born before 1924, initially a separate study (The Study of Assets and Health Dynamics Among the Oldest Old), baseline 1993
- Children of Depression (CODA) cohort, born 1924 to 1930, added to the study in 1998
- War Baby (WB) cohort, born 1942 to 1947, added to the study in 1998
- Early Baby Boomer (EBB) cohort, born 1948 to 1953, added to the study in 2004
- Mid Baby Boomer (MBB) cohort, born 1954 to 1959, added to the study in 2010

The data in this file include all six entry cohorts. The file only incorporates the core interviews. It does not include exit interview or any restricted data.

The RAND HRS Data is described in the [RAND HRS Data Documentation](#), which is included in this distribution. It describes the file in more detail and contains complete descriptions of the derived variables, including descriptions of how constructed, notes on cross-wave differences, and all raw HRS variables used.

RAND is grateful to NIA and SSA for past and continuing funding for the development of these data, to SSA for support to make them publicly available, and to the Office of

Research, Evaluation, and Statistics at SSA for providing important research direction in the design of these data files.

1.2. Confidentiality and Access Restrictions

The data described in this document are based on HRS public release files. Before using the data, you must have obtained permission from HRS by registering with them for downloading the public release files. The HRS website contains information on the process to register for access to HRS public release data (<https://ssl.isr.umich.edu/hrs>).

By registering with HRS you agree to the “Conditions of Use” governing access to the data. This agreement applies to the use of the RAND HRS Family data as well. There is NO RESTRICTED DATA on the RAND HRS Family data set.

2. Distribution Files

2.1. Description

The RAND HRS Data are distributed as a single file which includes the ten waves of the HRS or as nine separate files each containing one wave of data. The data contain respondents within the HRS, AHEAD, CODA, WB and EBB entry cohorts. The waves correspond to the following source data:

- Wave 1 is from 1992 V 1.01
- Wave 2 is from 1994 V1.0 data for the HRS cohort, and 1993 (V2.1) for the AHEAD cohort
- Wave 3 is from 1996 V4.0 data for the HRS cohort, and 1995 V2.0 for the AHEAD cohort
- Wave 4 is from 1998 V2.3
- Wave 5 is from 2000 V1.0)
- Wave 6 is from 2002 V2.0)
- Wave 7 is from 2004 Final V1.0 (October 2006)
- Wave 8 is from 2006 Final Release V2.0 (September 2010)
- Wave 9 is from 2008 Final Release V1.0 (November 2010)
- Wave 10 is from 2010 Final Release V1.0 (April 2013)
- Wave 11 is from 2012 Final Release V1.0 (December 2014)
- Wave 12 is from 2014 Early Release V1.0(October 2015)

The RAND HRS also uses cross-wave files to derive its variables, including the Tracker 2014 V1.0 released in December 2015, the Region and Mobility File V3.0 released in May 2010, and the Imputation of Cognitive Functioning Measures File 1992-2012.

The unit of observation is an individual. Each individual is uniquely identified by a household ID (hhid) and a person number (pn). We combined these variables into a single ID variable, hhidpn (HHIDPN: HHold ID + Person Number /Num), where hhidpn = 1000*hhid+pn. This file may be merged with other HRS data by hhidpn, or HHID and PN, separately. All RAND HRS Data files are sorted by HHIDPN.

The RAND HRS Data file is distributed in SAS, Stata and SPSS formats, as one file that includes all 11 waves.

The RAND HRS Data are distributed with the following:

- Documentation: an electronic version of [RAND HRS Data Documentation](#).
- Programs: source code of programs that were used to derive the RAND HRS Data files. All programs are written in SAS.
- SAS files: the data in SAS V9 format.
- Stata files: the data stored in longitudinal files are distributed in Stata 11 SE.
- SPSS files: the data in SPSS for Windows format.

This is version *P* of the RAND HRS Data. A variable called FileVer, with the single value “P”, identifies the version and appears on each file.

We suggest that you create a directory for these files and subdirectories for the pieces, for example:

C:\randhrs\doc	for this file
C:\randhrs\programs	for the programs
C:\randhrs\SASdata	for the SAS files
C:\randhrs\stata	for the Stata files
C:\randhrs\SPSS	for the SPSS files

2.2. Distribution files for Web Download

The files are zipped for downloading; you must unzip them to make them usable. They are available for download as an entire package or documentation only. There are four different format packages: SAS V9, Stata 11 SE, and SPSS.

The SAS files are available as one merged file, rndhrs_p.sas7bdat, containing all waves of data which is quite large but easiest to use. Some data conversion programs, such as StatTransfer, require that SAS files be unSAS-compressed. None of the SAS data files distributed are SAS-compressed. If you should need these files in a different format because you use SAS V6 on Unix, an earlier version of Stata, or use different statistical software, and do not have the tools to convert them, please contact us (randhrshelp@rand.org).

RAND HRS Data Distribution Files

All data files have 37495 observations. There are 11094 variables on the data files containing all waves. The SAS data file (randhrs_p.sas7bdat) requires about 3 GB, and the Stata file (randhrs_p.dta for Stata11 SE) about 700 MB. The SPSS data file (randhrs_p.sav) requires about 500 MB.

Distribution file name	Included files	Description
The complete package		
randPsas.zip	randhrsp.pdf	Codebook
	rndpgmp.zip	Programs
	randhrs_p.sas7bdat	SAS V9 data: all waves merged
	sasfmnts.sas7bdat	SAS formats for to create formats catalog
	rnd_Pdd.pdf	Data description (this file)
randPstata.zip	randhrs_P.pdf	Codebook
	rndpgmp.zip	Programs
	randhrs_p.dta	Stata 11 SE data: all waves merged
	rnd_Pdd.pdf	Data description (this file)
randPspss.zip	randhrs_P.pdf	Codebook
	rndpgmp.zip	Programs
	randhrs_p.sav	SPSS data: all waves merged
	rnd_Pdd.pdf	Data description (this file)
Documentation only		
rnddocp.zip	Randhrs_P.pdf	Codebook
rndpgmp.zip	all programs	Programs

3. Questions and Comments

Please let us know if you have any problems with or questions about the RAND HRS Data. Please direct your questions or comments about the RAND HRS Data to:

RANDHRSHELP@rand.org

RAND Center for the Study of Aging

www.rand.org/labor/aging

For more Information about data products please see: www.rand.org/labor/aging/dataproduct

1776 Main St.

P.O. Box 2138

Santa Monica, CA 90407-2138

Email: randhrshelp@rand.org