

The RAND HRS Data (Version K)

March 2011

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 2010, eleven HRS waves are available for study. The RAND HRS Data file is based on 1992, 1993, 1994, 1995, 1996, 1998, 2000, 2002, 2004, 2006 and 2008 final release data. The complete HRS includes five 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 1948-1953, added to the study in 2004

The data in this file include all five 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 RAND HRS Data are based on HRS public release files. Before using the data, you must have obtained permission from ISR by registering with them for downloading the public release files. By registering with ISR you agree to the “Conditions of Use” governing access to the data. This agreement applies to the use of the RAND HRS Data as well.

By receiving these data, which have been freely provided, you are agreeing to use them for research and statistical purposes only and make no effort to identify the respondents therein. In addition, you are in good faith agreeing to send HRS a copy of any publications you produce based on the data.

RESTRICTED DATA USERS, PLEASE NOTE: If you are using any HRS/AHEAD **restricted data** such as SSA data, you should check as to whether you may or may not merge them with the RAND HRS Data. If you intend to use the RAND HRS Data with restricted data please contact ([Cathy Leibowitz](#)) at ISR before doing so. Restricted data users are reminded that HRS/AHEAD must be informed of any data files used in conjunction with restricted data.

There are NO RESTRICTED DATA on this data set.

2. Distribution Files

2.1. Description

The RAND HRS Data are distributed as a single file which includes the nine 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)

The RAND HRS also uses cross-wave files to derive its variables, including the Tracker 2008 Final V1.0 released in October 2010, the Region and Mobility File V3.0 released in May 2010, and the Imputation of Cognitive Functioning Measures File 1992-2006 V1.0 released in September 2009.

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 9 waves and, and for Stata only, as 9 separate wave-specific files. To use the file that includes all 9 waves with Stata, one must have Stata Special Edition (SE). Intercooled versions of Stata limit the number of variables in its files, and can only use the 9 wave-specific files if all variables are kept. Stata 8+ intercooled can read Stata Special Edition (SE) files by selecting variables on the *use* command, as long as the total number of variables does not exceed 2047.

For example:

```
use hhidpn r1nhmliv r2nhmliv using "rndhrs_k.dta"
```

would select the respondent ID and 2 variables from the Stata SE File.

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 SE (Version 8+). The data split into wave-specific files are distributed in Stata Intercooled.
- SPSS files: the data in SPSS for Windows format.

This is version **K** of the RAND HRS Data. A variable called FileVer, with the single value “K”, 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 SE, Stata Intercooled, and SPSS.

The SAS files are available as one merged file, rndhrs_k.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 30,547 observations. Observations of individuals from the CODA and WB cohorts are missing all data for waves 1-3, and those from the EBB cohort are missing all data for waves 1-6. There are 7,992 variables on the data files containing all waves. The SAS data file (rndhrs_k.sas7bdat) requires about 2 GB, and the Stata file (rndhrs_k.dta for Stata-SE) about 375 MB. The by-wave Stata files are about 50 MB each. The SPSS data file (rndhrs_k.sav) requires about 325 MB.

Distribution file name	Included files	Description	
The complete package			
randKsas.zip	randhrsk.pdf	Codebook	
	rndpgmk.zip	Programs	
	rndhrs_k.sas7bdat	SAS V9 data: all waves merged	
	formats.sas7bcat	SAS format library for SAS users	
	sasfmts.sas7bdat	SAS formats for SPSS users	
	rnd_Kdd.pdf	Data description (this file)	
randKstataSE.zip	randhrsk.pdf	Codebook	
	rndpgmk.zip	Programs	
	rndhrs_k.dta	Stata 8 SE data: all waves merged	
	rnd_Kdd.pdf	Data description (this file)	
randKstataI.zip	randhrsk.pdf	Codebook	
	rndpgmk.zip	Programs	
	rndhrs1k.dta rndhrs2k.dta rndhrs3k.dta rndhrs4k.dta rndhrs5k.dta rndhrs6k.dta rndhrs7k.dta rndhrs8k.dta rndhrs9k.dta	Stata data, one file per wave, Stata 8 Intercooled . These include value labels assigned to variables where possible. Stata only allows labels for integer values. SAS special missing values are converted to Stata special missing values.	
	rnd_Kdd.pdf	Data description (this file)	
	randKspss.zip	randhrsk.pdf	Codebook
		rndpgmk.zip	Programs
		rndhrs_k.sav	SPSS data: all waves merged
		rnd_Kdd.pdf	Data description (this file)
	Documentation only		
rnddock.zip	randhrsk.pdf	Codebook	
rndpgmk.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