The RAND HRS Data File: A User-Friendly Version of the Health and Retirement Study

The Health and Retirement Study (HRS), a nationally representative sample of older Americans, is one of the leading resources for researchers and analysts studying policy topics on aging. Initially developed in 1992 by the Institute for Social Research at the University of Michigan, through support from the National Institute on Aging, the HRS collected health and socioeconomic information on respondents born between 1931 and 1941 through interviews at 2-year intervals. Over time, the HRS has added respondents from different age cohorts to develop a sample of more than 20,000 respondents over the age of 50, providing researchers with a rich collection of information on an aging America.

The complexity and sheer magnitude of the HRS files creates challenges for researchers who wish to use the panel file. The files contain thousands of variables. Matching respondent characteristics across interviews can be complicated, particularly for first-time users. Calculation of important economic variables such as income and wealth is complex and requires the use of imputation techniques for missing variables. Although the HRS is a valuable resource, a substantial investment of time is required to make use of the information in the study.

The RAND Corporation, with technical and financial support from the Social Security Administration and the National Institute on Aging, developed for public use an easy-to-use version of data from eight survey waves of the Health and Retirement Study called the RAND HRS.[1] The data incorporate selected information from the 1992, 1993, 1994, 1995, 1996, 1998, 2000, and 2002 interviews.[2] Although not as comprehensive as the HRS itself, the RAND HRS contains a subset of variables from the HRS that include demographics, income, health, employment, and wealth. Since the HRS identifiers exist on the RAND HRS, any variable in the main HRS data file can be matched to the RAND HRS.

This note was prepared by John R. Phillips, formerly of the Division of Policy Evaluation, Office of Research, Evaluation, and Statistics, Office of Policy, Social Security Administration. The author is currently a Health Scientist Administrator, National Institute on Aging, National Institutes of Health.

Features

The RAND HRS provides a useful collection of HRS variables for social science researchers to conduct analyses on topics about aging, particularly retirement. Features of the RAND HRS data include the following:

- A respondent-level design that matches all household and spousal data for a respondent to a single record for each survey wave;
- Consistent variable-naming conventions across all waves for easy matching of panel variables;
- All four birth cohorts of respondents, including the baseline HRS cohort (1931–1941), the AHEAD survey cohort (1890–1923), the Children of the Depression Age cohort (1924–1930), and the War Baby cohort (1942–1947);[3]
- Conceptual variables, which require substantial effort to create using the HRS data (examples include tenure on the longest job held by the respondent, the respondent's longest marriage, health indices, medical expenditures, and wages);
- Summary income and wealth data with imputations for missing data (the technique used to construct wealth and income data is used consistently across all panel waves of the data);
- Identifiers that allow users to match data from all HRS public and restricted-access data to the RAND HRS; and
- A comprehensive code book for the RAND HRS that describes the structure of the file, the construction and imputation of all variables, the HRS source variables used to construct all RAND HRS variables, and the means and frequencies of all variables.

Access to the Data

The RAND HRS data and code book can be downloaded from http://hrsonline.isr.umich.edu. All users must

register with the HRS before downloading the RAND HRS data.

For additional information regarding the RAND HRS, contact Patricia St. Clair at stclair@rand.org. For information regarding the HRS, contact the HRS staff at hrsquest@isr.umich.edu.

Access to Restricted Matched Data

Several useful restricted data sources can be matched to the RAND HRS to provide researchers with valuable information on the resources of older Americans. Social Security administrative data on earnings and benefits, data on employer pensions, and other data files are deemed "restricted"; researchers must apply for special permission to use them. These data files can be matched to the RAND HRS data only after receiving security clearance. Further, all use of restricted data must be done under secure conditions to protect the confidentiality of survey respondents. For more information on acquiring permission to match restricted data to the RAND HRS data, see http://hrsonline.isr.umich.edu/rda.

Notes

[1] Social Security Administration (SSA) provided support for the construction of the RAND HRS under Task Order 0440-00-40322, "Evaluation of the Effects of Changing Social Security Administration's Early Entitlement Age and the Normal Retirement Age," pursuant to Contract 0600-96-27335. SSA funding for the development and technical support of the RAND HRS is provided under a cooperative agreement with the University of Michigan Retirement Research Center. The National Institute on Aging provided support for the RAND HRS through grant P30-AG12815, "RAND Center for the Study of Aging."

[2] The current RAND HRS uses final versions of HRS data releases for all but the 2000 and 2002 interviews, which are preliminary releases. RAND is now incorporating the 2000 final version.

[3] The baseline HRS cohort was first interviewed in 1992, and follow-up interviews of the baseline cohort were conducted in 1994 and 1996. Interviews for the Asset and Health Dynamics Among the Oldest Old (AHEAD) cohort were fielded in 1993 and 1995 as a stand-alone survey. The HRS and AHEAD merged into one survey in 1998, and two more cohorts were added to the new HRS: the Children of the Depression Age (CODA) and the War Babies. All four cohorts were interviewed again in 2000 and 2002, with a follow-up scheduled for 2004. More information regarding the sample structure of the HRS can be found at http://hrsonline.isr.umich.edu.