

DATA PAPER

The Health and Retirement Study: A Public Data Resource for Research on Aging

Amanda Sonnega¹ and David R Weir²¹ Health and Retirement Study, Survey Research Center, Institute for Social Research, University of Michigan, USA² Director of the Health and Retirement Study, University of Michigan, USA

The Health and Retirement Study (HRS) is a nationally representative longitudinal survey of more than 37,000 individuals in 23,000 households over age 50 in the United States. Fielded biennially since 1992, it was established to provide a national resource for data on the changing health and economic circumstances associated with aging. HRS covers four broad topic areas—income and wealth; health, cognition, and use of health care services; work and retirement; and family connections. HRS data are also linked at the individual level to administrative records from Social Security and Medicare, Veteran’s Administration, the National Death Index, and employer-provided pension plan information. In 2006, data collection expanded to include biomarkers and genetics and greater depth in psychosocial well-being and social context. This blend of economic, health, and psychosocial information provides unprecedented potential to study increasingly complex questions about aging and retirement. HRS prioritizes rapid release of data while simultaneously protecting the confidentiality of respondents. Three categories of data—public, sensitive, and restricted—can be accessed through procedures described on the HRS website (hrsonline.isr.umich.edu).

Keywords: retirement; aging; national data; record linkage**Funding Statement:** HRS is grateful to the Behavioral and Social Research division of the National Institute on Aging of the National Institutes of Health for primary support of this project (U01 AG009740) and to the Social Security Administration for substantial co-funding.

(1) Overview

Introduction

In 1990, the U.S. Congress directed the National Institute on Aging (NIA) to create a new study, the Health and Retirement Study (HRS) [1] to provide scientific data for studying national-level social and policy changes that may affect individuals. The topics covered are broad and include resources for successful aging (e.g., economic, public, familial, physical, psychological, and cognitive); behaviors and choices (e.g., work, health behaviors, residence, transfers, use of programs); and events and transitions (e.g., health shocks, retirement, widowhood, institutionalization). HRS is now the largest nationally representative multidisciplinary panel study of Americans over age 50. The recent addition of biomarkers, genetics, and new psychosocial content make it the most comprehensive study of aging in the U.S. In addition, HRS has become the model and hub for a growing network of harmonized longitudinal aging studies around the world. HRS sister surveys currently include ELSA in England [2], TILDA in Ireland [3], 15 countries in the Survey of Health, Ageing and Retirement in Europe (SHARE) network and six surveys in Asia—IFLS in Indonesia, KLoSA in South Korea, CHARLS in China [4], LASI in India, HART in Thailand, and JSTAR in Japan. HRS is housed

within the Survey Research Center (SRC) at the Institute for Social Research (ISR) at the University of Michigan and works through a cooperative agreement with the NIA Division of Behavioral and Social Research (BSR).

Spatial coverage

The study takes place in the continental United States. The baseline interviews are conducted with community dwelling persons only. Participants who enter a nursing home after the baseline interview are retained in the sample and interviewed if possible.

Temporal coverage

The study began biennial data collection in 1992 and continues to the present.

Species

N/A

(2) Methods

Steps

The initial HRS cohort, recruited in 1992 for the study of retirement transitions, consisted of persons born 1931-41 (then aged 51-61) and their spouses of any age. A second

study, Asset and Health Dynamics Among the Oldest Old (AHEAD) was fielded the next year to capture an older birth cohort, those born 1890-1923. In 1998, the two studies merged, and, in order to make the sample fully representative of the older U.S. population, two new cohorts were enrolled, the Children of the Depression (CODA), born 1924-1930, and the War babies, born 1942-1947. The HRS now employs a steady state design, replenishing the sample every six years with younger cohorts to continue making it fully representative of the population over age 50. In 2004, Early Baby Boomers (EBB, born 1948-1953) were added, and in 2010, Mid Baby Boomers (MBB, born 1954-1959) were added. For respondents who are unwilling or unable to do an interview themselves, interviewers seek permission to use a proxy respondent. Proxies are usually a spouse or other family member. Use of proxies significantly improves sample retention [7] reducing a major source of non-random sample attrition in this survey of older adults. HRS also conducts follow-up interviews with next of kin following the death of a participant.

The main part of the survey, referred to as the core, takes place every two years, with the sample size ranging from about 22 to 25,000 at any given wave. Baseline interviews are conducted in person in participant's homes. Follow-up interviews are by phone unless the participant is over age 80 in which case follow-up interviews are conducted in person. Since 2006, a random half of the core sample gets an enhanced in-person interview at follow-up that includes physical measures (e.g., blood pressure, measured height and weight, timed walk), blood-based biomarkers, and genetics (from a saliva sample). A paper and pencil psychosocial questionnaire is left for participants to complete at their convenience and return by mail to the project office. The half-samples alternate waves so that the expanded content is available longitudinally every four years. HRS conducts supplemental studies on a variety of topics in the "off year" from the core survey. Samples are drawn from the core and range from 3,000-5,000 participants. Finally, HRS core data are linked at the individual level to Social Security earnings records, Medicare Claims, National Death Index, VA records, geographic information, and at the employer level to information on private pensions; sample sizes vary.

Sampling strategy

The HRS sample is based on a multi-stage area probability design involving geographic stratification and clustering and oversampling of certain demographic groups [5]. The primary and secondary stages of sampling involve sampling of 84 U.S. Metropolitan Statistical Areas and non-MSA counties. The third sampling stage involves a systematic selection of housing units within each of the sampled segments. The final stage in the multi-stage design is the selection of a financial unit within a sample housing unit. The design includes oversampling of African Americans and Hispanics. Weights are calculated and provided which account for the complex sample design as well as differential non-response. Initial response rates have declined over time (from 80 to 75%), following the

general national trend. Re-interview rates, however, have remained high (87-92%). HRS has been successful in recruiting and retaining minority participants [6].

Quality control

Based on the content areas and study design goals a data collection instrument is developed in computer-assisted interviewing (CAI), paper, and internet formats. Once the data collection instrument has been programmed, tested, and approved by the University of Michigan Institutional Review Board (IRB), it is ready to be transmitted to the interviewing team.

In most instances, actual interviewing is carried out by the Survey Research Operations (SRO) division of SRC. During the interviewing period, production (interim) datasets are transmitted to HRS for instrument validation and, if necessary, programming corrections. The production data sets are reviewed by HRS principal investigators who download encrypted files from a secure web site.

Raw data are delivered from SRO to the HRS staff offices. HRS staff generates database tables in a documentation database from the basic elements produced by deconstructing questionnaire or CAI meta-information (variable characteristics, question text, code frames, routing information, or respondent universe). HRS staff conducts a review of all fields in the raw data set(s) for possible respondent re-identification problems and assigns each variable to a distribution category. To build public-use data products, staff extracts data from raw files based on confidentiality review assessment; de-identified contents are considered suitable for public use. HRS staff generates complete documentation, in both ASCII and HTML formats, provided for each variable, including question text, code-frame, allowable ranges, universe and routing information, frequencies and univariate statistics, which are then provided on the public website download system.

Restricted data are prepared in the same format as public-use data. In order to preserve respondent confidentiality and to meet the specific conditions imposed on HRS by third-party data providers, data elements flagged by the confidentiality review as falling into the sensitive category can only be released under a special data agreement. Researchers may be eligible to receive HRS Restricted Datasets if they meet all of the following requirements:

- Affiliation with an institution with a DHHS-certified Human Subjects Review Process
- Current Receipt of Federal Research Funds
- Submission of a satisfactory research proposal
- Submission of an approved restricted data protection plan

Ethics

Collection and production of HRS data comply with the requirements of the University of Michigan's Institutional Review Board (IRB).

(3) Dataset description

Object name

Health and Retirement Study (HRS)

Data type

Secondary Data

Ontologies

N/A

Format names and versions

Most HRS data are provided in ASCII format, with fixed length records. Associated SAS, SPSS or STATA program statements are also provided that read the data into the analysis package of your choice. HRS provides several levels of files. Most files are respondent level files that contain data from questions that were asked of all respondents about themselves (or asked of a proxy about the respondent if the respondent was not able to give an interview). The files contain one record for each respondent or proxy who gave an interview in a given wave. Household level files contain data from family and financial questions asked of one respondent on behalf of the household. Sibling level files contain data on characteristics of the respondent's siblings. The sibling file contains one record for each sibling of a respondent. Other levels include helper level files, transfer-to-child-level files, and transfer-from-child files.

Creation dates

Since 1992, when the study began, HRS has produced public-release datasets approximately every two years for the core data and at various intervals for supplemental data collection projects.

Dataset creators

Together with the HRS faculty and staff at the University of Michigan, more than thirty researchers and professionals from other universities collaborate on the HRS study design and content. HRS operates through a cooperative agreement with the NIA Division of Behavioral and Social Research, which plays a pivotal role. In addition, the NIA Data Monitoring Committee (DMC) is an advisory group comprised of independent members of the academic research community as well as representatives of agencies interested in the study. All raw data are processed on-site at the University of Michigan, Institute for Social Research, Survey Research Center.

Language

English

Programming language

N/A

Licence

HRS requires new users to register and agree to several conditions of use detailed here; instructions for distribution to third parties are also outlined.

Accessibility criteria

HRS places a premium on early and open access to data while also implementing state-of-the-art data security measures to protect respondent confidentiality. Three categories

of data—public, sensitive, and restricted—can be accessed through the HRS website. Public data are available free to all registered users. Sensitive health data and restricted data (including linkages to Medicare, Social Security, Veteran's Administration, National Death Index, geographic information, and pension plan information) require submission of a separate data use agreement. Users wishing to link to HRS restricted data products must submit a restricted data application. Researchers wishing to use the HRS genetic data must first apply to the the NIH GWAS repository (dbGaP) for access to the genotyped data. Once access to dbGaP has been granted, researchers who wish to link to HRS phenotype measures not in dbGaP may apply for access to the HRS-dbGaP Cross-Reference File by submitting a Genetic Data Access Use Agreement (visit <http://hrsonline.isr.umich.edu/gwas> for more information).

Repository location

Most HRS data products are available free to registered users through the HRS website. Follow this link to register: <http://hrsonline.isr.umich.edu/index.php?p=reg>.

Publication date

Data are released on a rolling basis, usually within 3-5 months of the end of the field period.

(4) Reuse potential

Researchers at the RAND Corporation have created a user-friendly version of much of the HRS public data. Referred to as the RAND contribution and available through the HRS website, this version of the data is a good starting place for new users. Researchers at the University of Southern California have prepared cross-national data files for the HRS sister surveys, referred to as the Gateway to Global Aging and also available through the HRS website.

To encourage widespread use of the data, HRS staff conducts data use workshops in various locations throughout the year. An exhibit booth is also available at professional conferences with HRS staff available to help with questions about the data. Various resources for getting started with the data are available on the website, and an on-line helpdesk is offered for all users: hrsquestions@umich.edu. User outreach efforts have been successful with 14,700 registered users worldwide. Visit the HRS website (hrsonline.isr.umich.edu), especially under the documentation link, for more information on all of the topics addressed in this paper.

Acknowledgements

HRS gratefully acknowledges the contribution of the study participants who have given countless hours of their time to make this study what it is.

References

1. **Sonnega, A, Faul, J, Ofstedal, M B, Langa, K, Phillips, J and Weir, D** 2014 Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol* 43(2): 576–585, DOI: <http://dx.doi.org/10.1093/ije/dyu067>
2. **Steptoe, A, Breeze, E, Banks, J and Nazroo, J** 2013 Cohort profile: the English Longitudinal Study of Age-

- ing. *Int J Epidemiol* 42(6): 1640–1648, DOI: <http://dx.doi.org/10.1093/ije/dys168>
3. **Kearney, P M, Cronin, H and O'Reagan, C** 2011 Cohort profile: the Irish Longitudinal Study on Ageing. *Int J Epidemiol* 40: 877–884, DOI: <http://dx.doi.org/10.1093/ije/dyr116>
 4. **Zhao, Y, Hu, Y, Smith, J P, Strauss, J and Yang, G** 2011 Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). *Int J Epidemiol* 43(1): 61–68, DOI: <http://dx.doi.org/10.1093/ije/dys203>
 5. **Heeringa, S G and Connor, J H** 1995 Sample design and methods for the Health and Retirement Survey. Technical Report. *Statistical Design Group, Survey Research Center, University of Michigan*.
 6. **Ofstedal, M B and Weir, D R** 2011 Recruitment and retention of minority participants in the health and retirement study. *Gerontologist* 51: S8–S20, DOI: <http://dx.doi.org/10.1093/geront/gnq100>
 7. **Weir, D, Faul, J D and Langa, K** 2011 Proxy interviews and bias in the distribution of cognitive abilities due to non-response in longitudinal studies: A comparison of HRS and ELSA. *Longitud Life Course Stud* 2: 170–184, DOI: <http://dx.doi.org/10.14301/llcs.v2i2.116>

How to cite this article: Sonnega, A and Weir D R 2014 The Health and Retirement Study: A Public Data Resource for Research on Aging. *Open Health Data*, 2(1): e7, DOI: <http://dx.doi.org/10.5334/ohd.am>

Published: 16 October 2014

Copyright: © 2014 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.