

Appendix D: HRS Data Flow

1. Background

The Health and Retirement Study (HRS) is intended to provide data for researchers, policy analysts, and program planners who are making major policy decisions that affect retirement, health insurance, saving and economic well-being. The HRS is designed to supply researchers with the longitudinal data necessary to describe the trajectories of the economic, health and family status of Americans over age 50 and to test theories and estimate the parameters of dynamic behavioral models. The HRS is intended to facilitate inter-area research and encourage the creation of new conceptual frameworks by providing data in domains typically studied by separate disciplines. The mission of the HRS is to design, collect, and distribute high-quality data to the research community.

1A. Study Design

- National panel study designed to represent the United States population over age 50
- Initial sample of over 12,600 persons in 7,600 households; current sample of 26,000+ persons in 5,000+ households
- Over-samples (100%) of Hispanics, Blacks, and Florida residents
- Baseline: in-home, face-to-face in 1992 for the 1931-41 birth cohort (and their spouses, if married, regardless of age); in 1998 for 1924-30 (CODA: Children of the Depression Age) and 1942-47 (War Babies) birth cohorts; in 2004 for 1948-53 for Early Boomers.
- Incorporation of study of Asset and Health Dynamics Among the Oldest Old (AHEAD – pre- 1924 birth cohort) in 1998, with combined survey instrument and field data collection
- Follow-ups by telephone every second year, with proxy interviews after death
- A new cohort of early boomers born 1948-53 was added in 2004, with plans to enroll a new 6-year birth cohort each grant cycle.
- Subject population does not include participants under the age of 18.
- Some participants may be cognitively impaired; in these cases a proxy informant is sought.
- HRS does not follow subjects into prison or detention facilities.

1B. Content Areas

- Health and cognitive conditions and status
- Retirement plans and perspectives
- Attitudes, preferences, expectations, and subjective probabilities
- Family structure and transfers
- Employment status and job history
- Job demands and requirements
- Disability
- Demographic background
- Housing
- Income and net worth
- Health insurance and pension plans

1C. Administrative (Restricted) Data Links

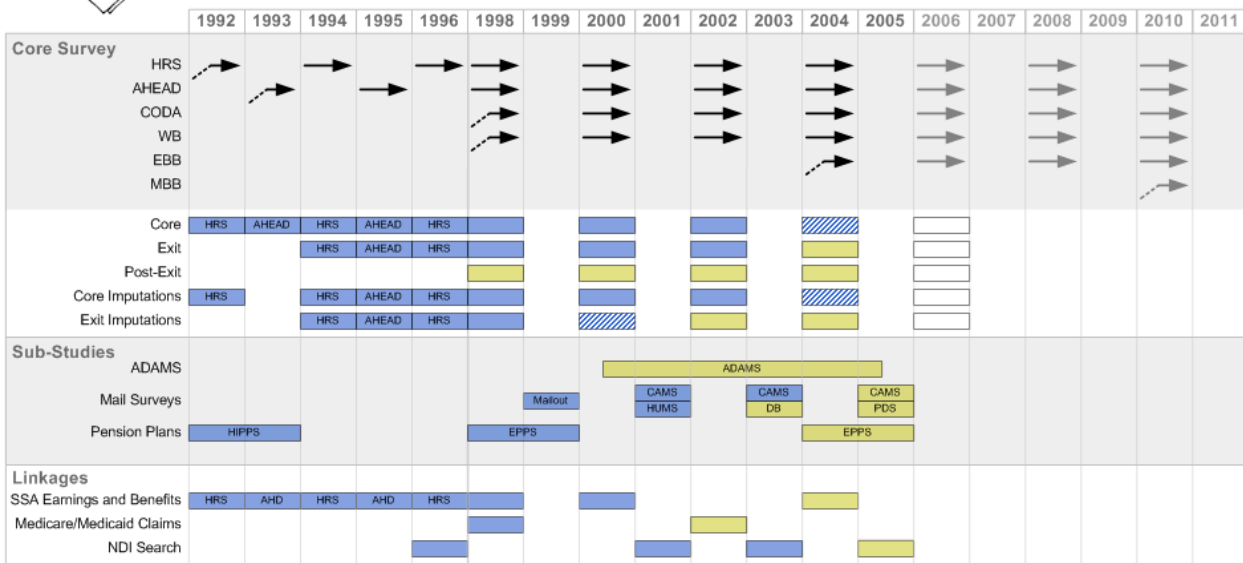
- Employer Pension Study (1993, 1999, 2005 - planned)

- National Death Index
- Social Security Administration earnings and benefits data
- W-2 self-employment data
- Medicare files (through 2000)

1D. Data Collection Plan



Health and Retirement Study: Data Collection Path



Cohort Key:
 HRS – Health & Retirement (1931-41)
 AHEAD – Aging & Health Dynamics (< 1923)
 CODA – Children of the Depression (1923-30)
 WB – War Baby (1942-47)
 EBB – Early Boomers (1948-53)
 MBB – Mid Boomers (1954-59)

Sub-Study Acronyms:
 ADAMS – Aging, Demographics & Memory
 HIPPS – Health Insurance & Pension Provider
 EPPS – Employer Pension Provider
 CAMS – Consumption & Activities
 HUMS – Human Capital & Educational Expenses
 DB – Diabetes
 PDS – Prescription Drug Study

Notes:

1. For product information, click on the appropriate arrow or box
2. Initial interview year for cohort: →
3. Use the **Question Concordance** as a tool for cross-referencing Core and Exit variables by content across time
4. Color Codes:
 - Yellow box: Data collection and/or processing
 - Blue box: Early release of data product
 - White box: Final release of data product
 - White box: Future data collection effort

2. Project Planning Activities

- Administration of Restricted Data contracts and data distribution
- Budgetary Planning
- Content area planning
- Methodology
- Monitoring Committee Support
- Providing User support (Help Desk)
- RFP responses
- Sponsor requests
- Steering Committee Support
- Study design

3. Instrument Development Activities

Based on the content areas and study design goals defined in the previous step, a data collection instrument is developed in either Computer-Assisted Interviewing (CAI) or Paper formats.

- **CAI.** Supports telephone interviewing, face-to-face interviewing and data entry. The system includes methods and utilities for conditional question routing, using multiple languages, coding, looking up data in external files, data manipulation, weighting data, Meta-data management and manipulation and for exporting data to external statistical packages or databases.
- **Paper.** Used for off-year and one-time surveys (i.e. off-year mail out surveys)

The following activities are involved in building a data collection instrument:

- Programming and testing data collection instruments
 - Blaise (2002 forward)
 - Surveycraft (1993-2000)
 - Paper and Pencil (1992, off-year mail surveys)
 - Internet (2000 forward)
- Maintenance of the testing database
- Instrument testing, including scenario generation and analysis of pre-test results
- Documentation creation (a paper questionnaire, a.k.a. 'Box and Arrow')¹
- All other data instrument creation activities

4. Data Collection

Once the data collection instrument has been programmed, tested, and approved by the IRB, it is ready to be transmitted to the interviewing team.² The delivery package includes:

- The CAI instrument
- Sample management files
- Preload data.

¹ A number of documentation methods have been developed by HRS to provide analysts with information on questions asked, allowable answers, routing instructions, and universe. These methods include human transformation of specifications into box and arrow format as well as computer generated equivalents using a variety of parsing techniques.

² In most instances, actual interviewing is carried out by the Survey Research Operations (SRO) division.

During interviewing period, production (interim) data sets 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.

5. Data Transformation

5A. Preliminary Work

- Receipt of final data from SRO
- Receipt of sample management files and other special files from SRO
- Metadata extraction: Generate 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, respondent universe)
- Confidentiality review: Review all fields in raw data set(s) for possible respondent re-identification problems; assign each variable to a distribution category: do not distribute, restricted use only, public masked/recoded, or public.

5B. Build Public-Use data products

- Extract data from raw files based on confidentiality review assessment; de-identified contents are considered suitable for public use.
- Generate 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.
- Prepare data distribution sets for installation in the Health and Retirement Study file distribution system.³

5C. Build Restricted data products

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 sensitive⁴ 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

6. Data Distribution

HRS data sets are available to the public from the User Registration and File Download Web site (<https://ssl.isr.umich.edu/hrs>). HRS data products are available without cost to researchers and analysts, however registration is required in order to download files. Offerings include:

³ <https://ssl.isr.umich.edu/hrs>/ Registration, contingent upon agreement to certain conditions of use, is required for access to data distributed from this site.

⁴ In this context, 'sensitive' denotes data elements that could be used to determine a respondent's actual identity. Review rules employed by HRS are based on HIPAA de-identification guidelines.

- *Final Release:* Data sets that have been cleaned, processed for ease of use, and supplemented by imputation, codebook, and user information files
- *Early Release:* Data sets that are made available a short time after the end of the field period, and have undergone minimal processing,
- *Cross-Wave Files:* Longitudinal datasets that link respondents, households and families across time.
- *Researcher Contributions:* Data sets provided by non-HRS researchers interested in sharing their work.
- *Imputation Files:* Amount variables with brackets, and variables that are integral components of aggregate variables such as household income, housing equity, and net worth
- *Documentation for Restricted Files*

7. Data Analysis

The Health and Retirement Study is intended to provide data for researchers, policy analysts, and program planners who are making major policy decisions that affect retirement, health insurance, saving and economic well-being.

- All HRS public data are made available to researchers via the HRS User Registration and File Download site (<https://ssl.isr.umich.edu/hrs>)
- Restricted data are available to qualified users who meet the requirements detailed on the Restricted Data Web site (<http://hrsonline.isr.umich.edu/rda>).
- Other services provided by HRS for researchers
 - The Help Desk answers questions dealing with all aspects of the study, via e-mail (<mailto:hrsquest@isr.umich.edu>) or online (<http://hrsonline.isr.umich.edu/intro/hrsquest.php>)
 - Question Concordance (<http://hrsweb.isr.umich.edu/concord>): a searchable database of biennial interview questions; used to trace cross-wave responses
 - Online Bibliography (<http://hrsweb.isr.umich.edu/biblio/index.html>): a searchable database of books, journal articles, monographs and working papers containing information derived from HRS data.
- Researcher input to project planning:
 - HRS Steering Committee: Members provide advice to HRS on tasks to be performed by the expert working groups, monitor progress of the study, and make recommendations about sample design, survey content and operational procedures.
 - National Institute of Aging -- HRS Monitoring Committee: an advisory group comprised of independent members of the academic research community as well as representatives of agencies interested in the study. The structure of the Monitoring Committee allows and encourages the coordination of federal statistical operations, which increases the value of the study to both federal and research communities. This committee provides advice and recommendations to the NIA Program Officer.