



LABOR AND POPULATION

***Using the RAND Versions  
of HRS Data***

**RAND Center for the Study of Aging**

# ***RAND HRS Data Files***

- **Supported by the National Institute on Aging (NIA) and the Social Security Administration (SSA)**
- **Developed from the Health and Retirement Study**
- **Developed by the RAND Center for the Study of Aging**

# *Purpose*

**To facilitate empirical research on aging issues by making the HRS data easier to use.**

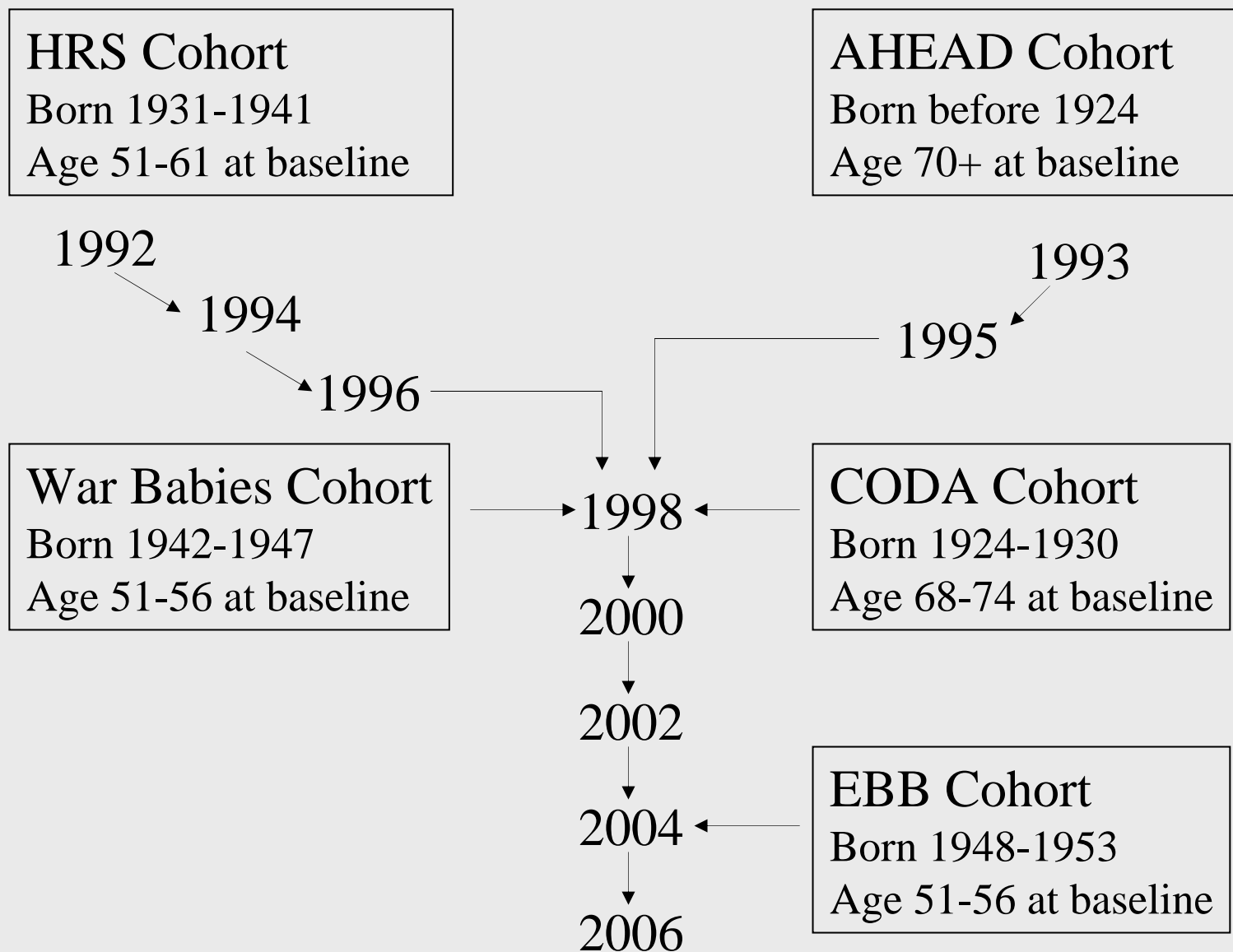
# *Overviews*

- **The Health and Retirement Study**
- **RAND-enhanced “fat” files**
- **RAND HRS Data file**
- **Research Uses**
- **Current Status and Work in Progress**
- **Using the RAND Data Products**

# ***HRS Background***

- **Longitudinal survey conducted every two years**
- **Covers a broad range of topics, including health, income, assets, employment, retirement, insurance, and family structure**
- **Collected by the Institute for Social Research at University of Michigan, funded mainly by NIA**
- **Nationally representative of the older population**

# The HRS Sample: Birth year entry cohorts and their spouses



## ***RAND-enhanced Fat Files***

- **Supported by NIA**
- **One for each HRS/AHEAD wave**
- **All raw HRS variables included, merged properly to the respondent level**
- **Added variables to clarify to whom the data apply**
- **Unusual or discrepant data documented**
- **Easily merged with other waves**
- **Income and Wealth imputations**

# ***Fat File Structure Example***

- If you were to look at the HRS 2006 data, you will find a raw variable named “KQ015”, which is the amount that the respondent made from self-employment last calendar year.
- This same information for the spouse is actually in the raw variable “KQ040”.

	KQ015	KQ040	KQR015	KQP015
JACK	300	1500	300	1500
JILL	300	1500	1500	300

# ***RAND HRS Data File***

- **SSA conceived of and funded further HRS data development**
- **File includes lots of derived variables**
- **All entry cohorts: HRS, AHEAD, CODA, and WB, EBB Waves 1-8**
- **Based on public use data only, i.e., no restricted data**

# ***RAND HRS Data File***

- **Data: all waves, respondent level observations**
  - **Consistently named derived variables**
  - **Comparable across waves**
  - **Derives variables longitudinally when needed**
  - **Includes analogous spouse variables**
- **Documentation**
  - **How variables constructed**
  - **All raw HRS variables listed**
  - **Cross wave differences noted**
- **SAS Programs**

# What's in which wave

<b>Wave</b>	<b>HRS Data Used</b>
<b>1</b>	<b>HRS 1992</b>
<b>2</b>	<b>Ahead 1993 and HRS 1994</b>
<b>3</b>	<b>Ahead 1995 and HRS 1996</b>
<b>4</b>	<b>HRS 1998</b>
<b>5</b>	<b>HRS 2000</b>
<b>6</b>	<b>HRS 2002</b>
<b>7</b>	<b>HRS 2004</b>
<b>8</b>	<b>HRS 2006</b>

# *Derived Variable Examples*

- **Years of tenure at current job**
  - Respondent only asked about job start date at first interview (if they are working).
  - At subsequent interviews, changes in work status trigger questions about stop date of job that ended.
  - Thus, we carry the amount of time in the job forward.

# *Derived Variable Examples*

- **“When did you last work?”**
  - **Many places to look**
    - **If not working at the first interview, R is asked about the most recent job and up to 3 other past jobs**
    - **If working at the previous interview but not now, R is asked when stopped last job**
    - **If employment status is retired, disabled, or unemployed, R is asked when status began, i.e., when stopped working**
    - **Derivation uses all these sources as needed**

Sample of Derived Variable  
Documentation from RAND HRS  
Codebook

## Self-report of health

Cross Section Files: 1 to 8

Wave	Variable	Label	Type
1	R1SHLT	R1SHLT:W1 Self-report of health	Categ
2	R2SHLT	R2SHLT:W2 Self-report of health	Categ
3	R3SHLT	R3SHLT:W3 Self-report of health	Categ
4	R4SHLT	R4SHLT:W4 Self-report of health	Categ
5	R5SHLT	R5SHLT:W5 Self-report of health	Categ
6	R6SHLT	R6SHLT:W6 Self-report of health	Categ
7	R7SHLT	R7SHLT:W7 Self-report of health	Categ
8	R8SHLT	R8SHLT:W8 Self-report of health	Categ
1	S1SHLT	S1SHLT:W1 Self-report of health	Categ
2	S2SHLT	S2SHLT:W2 Self-report of health	Categ
3	S3SHLT	S3SHLT:W3 Self-report of health	Categ
4	S4SHLT	S4SHLT:W4 Self-report of health	Categ
5	S5SHLT	S5SHLT:W5 Self-report of health	Categ
6	S6SHLT	S6SHLT:W6 Self-report of health	Categ
7	S7SHLT	S7SHLT:W7 Self-report of health	Categ
8	S8SHLT	S8SHLT:W8 Self-report of health	Categ

## Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R1SHLT	12652	2.584	1.205	1.000	5.000
R2SHLT	19632	2.813	1.192	1.000	5.000
R3SHLT	17984	2.790	1.176	1.000	5.000
R4SHLT	21378	2.922	1.177	1.000	5.000
R5SHLT	19571	2.837	1.156	1.000	5.000
R6SHLT	18156	2.877	1.133	1.000	5.000
R7SHLT	20112	2.883	1.142	1.000	5.000
R8SHLT	18444	2.884	1.132	1.000	5.000
S1SHLT	9900	2.513	1.173	1.000	5.000
S2SHLT	13084	2.705	1.163	1.000	5.000
S3SHLT	11911	2.689	1.149	1.000	5.000
S4SHLT	13973	2.819	1.151	1.000	5.000
S5SHLT	12726	2.732	1.129	1.000	5.000
S6SHLT	11639	2.751	1.104	1.000	5.000
S7SHLT	12962	2.767	1.119	1.000	5.000
S8SHLT	11723	2.760	1.111	1.000	5.000

## Categorical Variable Codes

Value-----	R1SHLT	R2SHLT	R3SHLT	R4SHLT	R5SHLT	R6SHLT	R7SHLT	R8SHLT	
.D=DK/NA		8	3	6	7	8	13	23	
.M=Oth missing			2					1	
.R=RF		2	2		1	3	4	1	
1. Excellent		2807	2982	2664	2633	2473	2050	2032	
2. Very good		3481	5241	5078	5467	5652	5081	5476	
3. Good		3544	5812	5262	6541	5903	5739	6280	
4. Fair		1807	3660	3325	4400	3681	3616	4135	
5. Poor		1013	1937	1655	2337	1862	1670	1858	
Value-----		S1SHLT	S2SHLT	S3SHLT	S4SHLT	S5SHLT	S6SHLT	S7SHLT	S8SHLT
.D=DK/NA			3	2	5	4	2	7	12

	1	2	3	4	5	6	7	8	9
.M=Oth missing									
.R=RF			1						3
.U=Unmar		2373	5970	5658	6869	6538	6306	6777	6418
.V=Sp NR		379	584	418	537	311	220	380	316
1. Excellent		2293	2166	1926	1871	1769	1499	1718	1502
2. Very good		2848	3819	3618	3887	3984	3581	3857	3627
3. Good		2809	3844	3487	4360	3842	3700	4080	3612
4. Fair		1287	2225	1990	2606	2147	2037	2339	2144
5. Poor		663	1030	890	1249	984	822	968	838

## How Constructed:

RwSHLT is the respondent's self-reported general health status. Codes range from 1 for Excellent to 5 for Poor. SwSHLT is the respondent's spouse or partner's self-reported general health status.

RwSHLT is assigned the value of the raw variable except that missing values for don't know, refused, and other missings are recoded to .D, .R, and .M, respectively.

RwSHLT and SwSHLT are used in construction of a change in health variable RwSHLTC. Please see "Change in Health" for a description of these measures.

The SwSHLT variables are taken from the Wave 'w' spouse's self-reported RWSHLT variables.

## Cross Wave Differences in Original HRS Data

In Wave 1 values for self-reported health status are imputed by HRS if missing. These imputations are used. From Wave 2 forward, values are not imputed by HRS.

## HRS Variables Used

HRS 1992:	
V301	B1:RATE CURRENT HEAL:IMP
AHEAD 1993:	
B204	B1. RATE HEALTH
HRS 1994:	
W301	B1.RATE CURRENT HEALTH
AHEAD 1995:	
D769	B1. RATE HEALTH
HRS 1996:	
E769	B1. RATE HEALTH
HRS 1998:	
F1097	B1. RATE HEALTH
HRS 2000:	
G1226	B1. RATE HEALTH
HRS 2002:	
HC001	RATE HEALTH
HRS 2004:	
JC001	RATE HEALTH
HRS 2006:	
KC001	RATE HEALTH

***Part of HRS 98 V1 RAND-  
Enhanced Fat File  
Documentation***

## HRS 1998: Notes Summary

### Specific Changes and Additions to the Public Release HRS 1998 Data.

HRS 1998 includes most of the public release data. "Other Person" files containing data on children, siblings, household members, helpers, and transfers between HRS respondents and their children, with observations for each other person, are distributed by ISR but have not yet been included in the RAND Enhanced Flat Files.

HRS Data Set	
Type of Change / Addition	HRS 1998 (Final Version 1)
<u>Generally helpful variables</u>	In <i>h98fla</i> SAS file
<u>Financial Variables Added, including assignment to "self" and "spouse" variables for the individual</u>	[ <b>Note:</b> the 1998 flat file is distributed as one rather large data set containing all cohorts, or as four smaller separate files, one for each cohort:
<u>Family Variables Added, including assignment to "self" and "spouse" variables for the individual</u>	<ul style="list-style-type: none"> <li>• h98fla, all cohorts included</li> <li>• h98fla_a for AHEAD respondents</li> <li>• h98fla_c for CODA respondents</li> <li>• h98fla_h for HRS respondents</li> <li>• h98fla_w for War Babies respondents</li> </ul>
<u>Summary of Notes on Skip Patterns and Problems</u>	To get the complete sample, one would use either h98fla OR the combination of h98fla_a, h98fla_c, h98fla_h, and h98fla_w.]
<u>RAND Imputations for Income and Assets</u>	In <i>income98</i> and <i>wealth98</i> SAS files

### Notes about generally helpful variables

- Tracker not completely updated for core
- Illogical interview dates
- Marital Status Variable Derivations
- Gender Discrepancies
- Birth Year Discrepancies
- Race and Whether Hispanic Discrepancies

### Notes about Financial Variables

- Income Variables Assigned to Self/Spouse-Partner Variables (FR/FP)
- Who worked for pay last year
- Who receives transfer, IRA/pension, trust income
- Whose parents an inheritance is from
- Derived categorical bracket variables
- Random entry points for unfolding bracket questions
- Preloads for random entry points for unfolding bracket questions
- How much income last year (rent, own); business/farm, dividends/interest
- Income last calendar year derived from monthly income questions-veteran benefits, social security, SSI, pensions/annuities, food stamps
- Income last calendar year variables recoded for missing values
- Missing income amounts filled from unfolding bracket questions
- Warnings about skip patterns or data
  - Notes on unfolding bracket questions for self-employment and wages
  - Notes on unfolding bracket questions for other earned income
  - Possible Apparent Discrepancies from Comment Coding

### Notes about Family Variables

- how much help R and Spouse gave to own parents and in-laws
- shortened names for variables already 8-characters, to accommodate "FR" prefix
- number of variable occurrences different for FamR's own family vs in-laws
- relationship codes adjusted for non-FamR
- opinions solicited from FamR
- Family Variable Assignments.

# Easing into the Analysis of HRS Data

## **RAND HRS Data**

- Basic means and frequencies
- Codebook:
  - Cross wave differences
  - Variable lists
- Programs:
  - “borrow” code
  - SAS Macros

## **RAND Fat Files**

- Most raw HRS variables
- Added variables
- Notes and warnings

## **HRS Website, Codebooks, and Questionnaires**

- Documents all raw HRS variables
- Provides question text
- Documents instrument skip patterns
- Provides concordance help

## **Data plan for your analysis**

- Basic “lay of the land”
- Potential problems
- Data sources:
  - RAND HRS Data
  - “raw” variables from Fat Files
  - HRS public release files

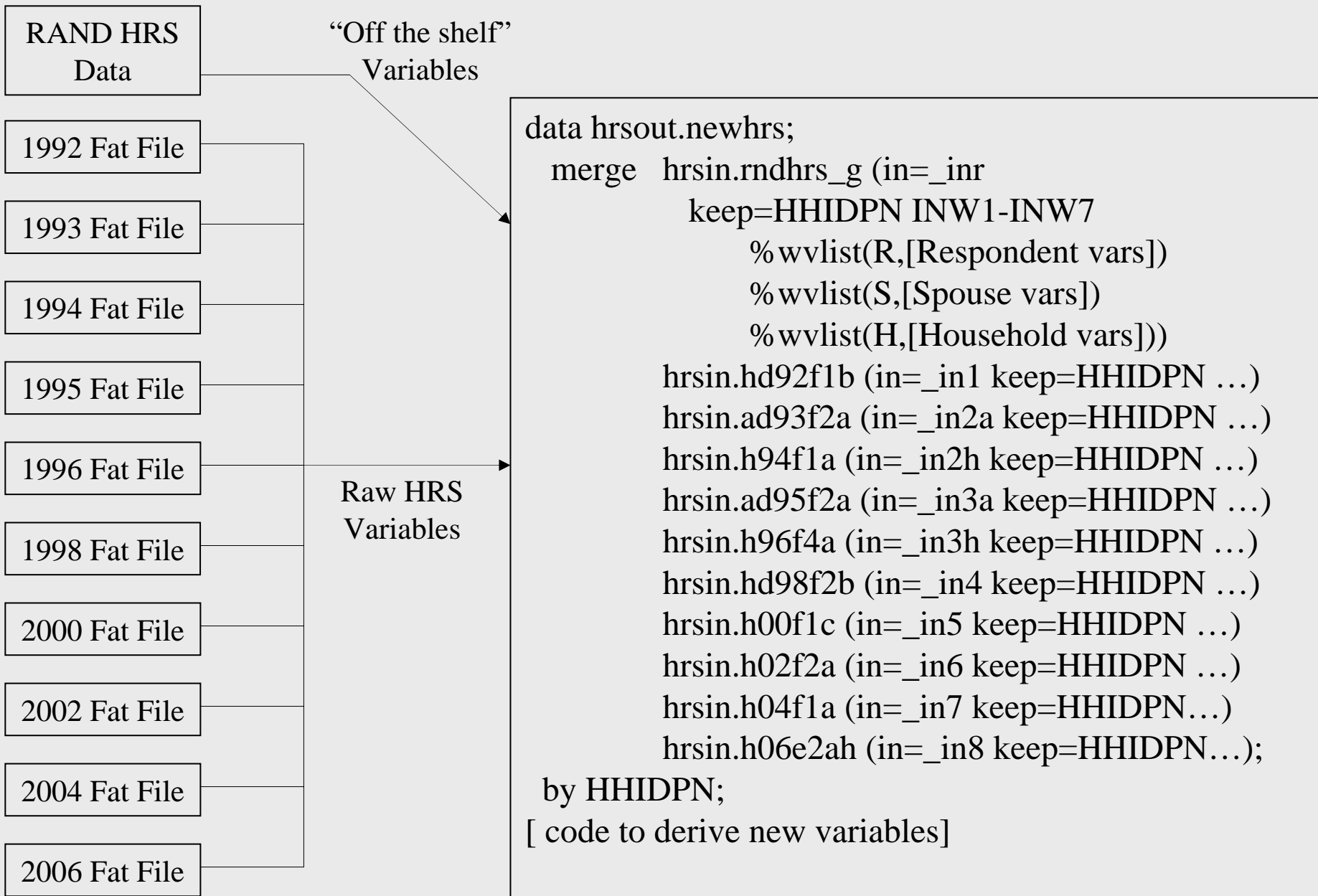
# *Research Uses*

- **Health**

- Health behaviors, such as smoking, drinking, and preventative behaviors
- Medical care utilization and imputed out of pocket and total medical expenditures
  - Questions answered by each member of the household

- **Financial and Housing Wealth**

- Various totals, including total wealth, net housing wealth, and total non-housing wealth
- Debt
  - Questions answered by financial respondent for the entire household



# *Current Status*

- **RAND HRS Data file for HRS, AHEAD, CODA, WB, and EBB entry cohorts, Waves 1-8, Version H**
  - **Final versions of HRS 1992-2004**
  - **Early release version of 2006 (V2.0)**
  - **Tracker 2006 V1.0 (October 2007)**
- **RAND-enhanced “fat” files for 1992, 1993, 1994, 1995, 1996, 1998, 2000, 2002, 2004, and 2006.**

# *Work in Progress*

- **RAND HRS Data file - Version I**
  - Awaiting final release of HRS 2006
  - Considering additional variables, making adjustments to outlier methodology, etc.
- **Continue improvements to data**
- **Work supported by NIA and SSA**

# *Using the RAND HRS Data Products*

- **Overview**
  - **File characteristics**
  - **IDs**
- **RAND HRS Data File**
  - **Downloading and unpacking**
  - **Using the codebook**
  - **Running tables and summary statistics**
- **RAND-enhanced Fat Files**
  - **Requesting and unpacking**
  - **Merging**

## ***Characteristics of the files***

- **RAND HRS Data File:**
  - Longitudinal (or easily made so)
  - Respondent level, sorted by HHIDPN / RAHHIDPN
  - Variables consistently named across waves
- **RAND Fat Files**
  - Wave-specific, one file per interview year
  - Respondent level, sorted by HHIDPN / RAHHIDPN / HHID&PN
  - Variables named as in Public Release data
- **Easily merged together**
- **SAS, SPSS, and Stata (Stata SE only for Fat Files)**

## ***Respondent Identification***

<b>Data Set</b>	<b>Variables</b>	<b>Example</b>
<b>HRS Public Release</b>	HHID [char 6] PN [char 3]	“054321” “010”
<b>RAND HRS Data File</b>	HHIDPN [num] RAHHIDPN [char 9]	54321010 “054321010”
<b>RAND Fat Files</b>	HHID [char 6] / HHIDN [num] PN [char 3] / PNN [num] HHIDPN [num] RAHHIDPN [char 9]	“054321” / 54321 “010” / 10 54321010 “054321010”

## *Household Identification*

<b>Data Set</b>	<b>Variables</b>	<b>Example</b>
<b>HRS Public Release</b>	HHID [char 6] xSUBHH [char 1]	“054321” “0”
<b>RAND HRS Data File</b>	HwHHID [num]	543210
<b>RAND Fat Files</b>	HHID [char 6] / HHIDN [num] xSUBHH [char 1] xSUBHHN [num] xHHID [char 7] xHHIDN = HwHHID [num]	“054321” / 54321 “0” 0 “0543210” 543210

# ***Making compatible IDs from HRS public release variables***

- **SAS:**
  - RAHHIDPN = HHID || PN;
  - For 2000: H5HHID = HHID\*10+GSUBHH;
- **Stata:**
  - gen str9 rahhidpn=hhid + pn;
  - For 2000: gen long h5hhid=real(hhid)\*10+real(gsubhh);
- **SPSS:**
  - String RAHHIDPN (A9).
  - compute RAHHIDPN = concat(HHID,PN).
  - For 2000: compute H5HHID=number(HHID,F6.0)\*10+number(gsubhh,F1).
- **OR**
  - add RAHHIDPN and HwHHID from a fat file by merging on HHID and PN

# *Overlap cases*

- **HRS-AHEAD overlaps**
  - **Interviewed in 1992 as HRS households, with HRS IDs**
  - **Given to AHEAD 1993, and continue to be followed as AHEAD households, with AHEAD IDs**
  - **Linked together in RAND HRS**
    - **RAOVLAP=1**
    - **HHIDPN, RAHHIDPN = AHEAD ID**
    - **RAOHRSID = HRS ID**
    - **RAOAHDID = AHEAD ID**
  - **Assigned AHEAD IDs in hd92f1b Fat File and HRS IDs in h92f1b Fat File**

## *More Overlap cases*

- **AHEAD-AHEAD overlaps**
  - **Married into a different AHEAD household**
  - **Linked longitudinally in RAND HRS**
    - **RAOVLAP=2**
    - **RAOVRAYR = 1995 or 1998, the last year with original AHEAD ID**
    - **HHIDPN, RAHHIDPN = most recent AHEAD ID**
    - **RAOAHDID = original AHEAD ID**
  - **Assigned most recent AHEAD IDs in Fat Files ad93f2a, ad95f2a, and hd98f2b, and original AHEAD IDs in a93f2a, a95f2a, and h98f2b.**

## *Setting up*

- **RAND HRS Data File - the pieces**
  - **Data file (s)**
    - **SAS V8**
      - **Use formats.sas7bdat for value labels**
    - **Stata V8 or SE**
      - **Intercooled must use single-wave files**
      - **SE may use longitudinal file**
    - **SPSS**
  - **Codebook**
  - **Programs**

## ***Downloading the RAND HRS Data File***

- **HRS Web Site, Public File Download Area**  
**Login:**

- <https://ssl.isr.umich.edu/hrs/login.php>
- **Datasets and Files link**
- **RAND Contributed Files (lower right)**
- **RAND HRS Data File link**
- **Big package in one download**

# ***RAND HRS Data File – the codebook***

- **Introduction**
  - **Background and overview**
  - **Technical information and help**
- **Data Codebook**
  - **Table of contents / bookmarks**
  - **What's on a page**
    - **Statistics**
    - **How constructed and raw variables**
    - **Cross-wave differences**

## ***RAND HRS Data File – Wave Status***

- **Who responded: INWw flags**
  - **INWw=1** if responded in wave “w”
- **Mortality status in wave “w”: RwlWSTAT**
- **SwHHIDPN: Spouse HHIDPN in wave “w”**
  - **If not married, spouse variables = .U**
  - **If married but spouse didn’t respond, spouse variables = .V**
- **If R is non-response in a given wave, most variables set to plain missing (.)**

## ***Getting the RAND Fat Files***

- **Mailed on a CD upon request**
- **Data on the CD is encrypted**
- **Pass-phrase is available at the**
  - **HRS web site in the Public File Download Area**
  - **under RAND Contributions**
  - **RAND HRS Distribution CD Key**

# *Questions and Comments*

**RAND Center for the Study of Aging**

**[www.rand.org/labor/aging](http://www.rand.org/labor/aging)**

**RAND HRS Help**

**Email: [RANDHRSHelp@rand.org](mailto:RANDHRSHelp@rand.org)**