

How to Use the Health and Retirement Study

A simple researcher's primer

Part 2

Ryan D. Edwards

redwards@qc.cuny.edu

Queens College, City University of New York & NBER
Visiting UC Berkeley Demography 2012-2013

March 22, 2013

Broad outline

- The Health and Retirement Study:
 - What is it?
 - What's in it?
 - How do you get access to it?
 - How do you work it?
-
- The diagram illustrates the broad outline of the HRS study. It starts with a main bullet point about the HRS. Below it, two pairs of questions are listed. The first pair ('What is it?' and 'What's in it?') is grouped by a large black brace on the right, labeled 'Tell'. The second pair ('How do you get access to it?' and 'How do you work it?') is also grouped by a large black brace on the right, labeled 'Show' in blue text.

HRS: How do I access it?

- There is no online analysis tool, you need to download and load the data into a statistical program of your choice
- Datafiles are available to registered users of the HRS website at
<http://hrsonline.isr.umich.edu>
 - All the HRS core and exit interview files
 - The tracker file, which measures individuals longitudinally
 - The RAND files, which draw from these
 - Documentation on these and on the restricted and sensitive files
- Registering is straightforward

HRS: How do I work it?

- Where to start? (My suggestion: With the RAND file)
- Advice about deciphering original HRS data releases
- Syntax of merging multiple datasets
- Reshaping the longitudinal RAND file from wide to long
- Hands-on with an example using the RAND file

HRS is complex, which has upsides and downsides

- HRS is so rich that there's a good chance that your ideal question has been asked sometime of somebody
- The challenges are:
 - Finding out when and how it was asked, and of whom. Use the online HRS “Question Concordance” fillable search page on the HRS website.
 - Linking it to individuals' or to households' other variables; in particular, bringing data forward or backward in the panel
 - Making sure that the “don't know” or “refused” are coded reasonably; they typically are 998 or 999 in the primary data distributions
- The RAND file is an excellent place to start, and I recommend that. So let's look at the RAND file and discuss merging multiple datasets

If you need to access the original HRS data releases

- Begin with the [HRS “Question Concordance” fillable search page on the HRS website](#) to identify and locate your prey
- If that leads to no single answers, you can try looking at the original datafiles themselves
- Sometimes the lists of variables in each dataset are so long and detailed as to be seemingly useless
- I recommend that you start with the Stata dictionary files (.dct), which list one variable per line, and search for keywords
- You can also keyword-search the concatenated codebook text file (e.g., h2010cb.txt), or the questionnaire PDFs
- Beware of numeric codes for “Don’t know” and “Refused” and others

Merging mechanics in Stata with an original data file

- `clear`
- `set more off`
- `infile using H08B_R.dct`
- `save H08B_R.dta, replace`
- `set more on`
- `gen hhidpn = HHID + PN`
- `destring hhidpn, replace`
- `merge 1:1 hhidpn using rndhrs_1.dta`

First step: Read in
original data with
`infile`

- If you want to join data to individuals, the linking variable is `hhidpn`
- In most HRS Stata .dct files, HHID and PN are read as strings
- Concatenate strings with “+” then destring because the RAND file uses numeric (long) `hhidpn`
- Beware of precision issues with `hhidpn`, double check your work

Alternative merging mechanics in Stata

- . destring HHID, replace
- . destring PN, replace
- . gen long hhidpn = HHID*1000 + PN
- . merge 1:1 hhidpn using rndhrs_1.dta
- . merge 1:1 HHID PN using H08B_R.dta
- Much of the RAND documentation implies they use this top formulation
- Beware precision issues
- I prefer the string concatenation shown on the previous slide
- You can also merge by both HHID and PN if you want
- But the RAND file only has hhidpn, so that requires an extra step

Key variables and naming conventions in the RAND file

hhidpn	Individual identifier	rKiwstat	Respondent's interview status in wave K (mortality and more)
hacohort	HRS birth cohort assignment (e.g., AHEAD, HRS, CODA)	inwK	Whether respondent is in wave K
rabyear	Respondent's birth year	sKgender	Sex of spouse of R in wave K
ragender	Respondent's sex		
raedyrs	Respondent's education in years	sKedyrs	Education in years for spouse of R in wave K

Reshaping mechanics in Stata

	hhidpn	r1shlt	r2shlt	r3shlt
1	1010	2	3	4
2	2010	3	3	2



```
. clear  
. use rndhrs_1.dta  
. keep hhidpn r*shlt  
. reshape long r@shlt, i(hhidpn)  
j(wave)
```

(no line break in the reshape command)

- RAND HRS file comes in a “wide” format (at left)
- Each individual gets a single row
- For longitudinal analysis, often want “long” format: each row a person-wave

	hhidpn	rshlt	wave
1	1010	2	1
2	1010	3	2
3	1010	4	3
4	2010	3	1
5	2010	3	2
6	2010	2	3

Application: Health status and smoking (1/4)

- Does smoking worsen health? Only a social scientist would question whether it in fact does
- (Aside: Yes, I believe smoking is bad for you. Kids, don't smoke.)
- Let's use this question to explore the RAND HRS file. We want to examine:

$$y_i = \alpha + \beta x_i + X_i B + \varepsilon_i \quad (1)$$

where x is smoking, y is self-reported health, and X contains controls

- In HRS and many other surveys, self-reported health is “inverted,” with 1 = excellent, 2 = very good, 3 = good, 4 = fair, 5 = poor
- Smoking is bad for health if $\partial y / \partial x > 0$
- With the HRS, we can estimate equation (1) by (i) ordinary least squares, by (ii) instrumental variables if we link in cigarette prices with geocodes, and by (iii) panel estimators

Application: Health status and smoking (2/4)

$$y_i = \alpha + \beta x_i + X_i B + \varepsilon_i \quad (1)$$

- Ordinary least squares estimation is likely to be biased by omitted variables when we don't have the right X's
 - There are probably unmeasurables that prompt people to have bad health AND to smoke
 - (And maybe health influences the decision to smoke)
- A good instrumental variable like cigarette prices, which are external to the individual and health, could reveal a causal pathway of smoking
- Panel fixed-effects can also shed light on equation (1) but do not solve all problems of causality; it merely reveals differences in x's associated with differences in y's

Application: Health status and smoking (3/4)

hhidpn	Individual identifier	rKagey_m	Respondent's age in years at middle of interview in wave K
ragender	Respondent's sex (2 categories)	rKshlt	Respondent's self-reported health status (1,2,3,4,5, 1 = excellent, 5 = poor)
raracem	Respondent's race (3 categories: white, black, other)	rKsmoken	Respondent self-reports smoking cigarettes now
rahispan	Respondent's Hispanic status	raedyrs	Respondent's years of education

Application: Health status and smoking (4/4)

- . use "/Users/redwards/Data/HRS/RAND/randLstataSE/rndhrs_1.dta"
- . keep hhidpn r*shlt r*smoken ragender raracem rahispan r*agey_m raedyrs
- . reg r10shlt r10smoken i.ragender i.raracem i.rahispan raedyrs r10agey_m, vce(robust)
- . drop raedyrs
- . reshape long r@shlt r@smoken r@agey_m, i(hhidpn ragender raracem rahispan) j(wave)
- . reg rshlt rsmoken i.ragender i.raracem i.rahispan ragey_m i.wave, vce(robust)
- . tsset hhidpn wave
- . xtreg rshlt rsmoken i.ragender i.raracem i.rahispan ragey_m i.wave, fe vce(robust)

Some references with hyperlinks

James Banks, Alastair Muriel, and James P. Smith (2010) “Attrition and Health in Ageing Studies: Evidence from ELSA and HRS” *RAND Labor and Population working paper*

David Weir (2010) “Socio-economic Status and Mortality: Perceptions and Outcomes,” UCLA seminar paper prepared for PAA 2010.

HRS “Question Concordance” fillable search page on the HRS website.

HRS “Sample Evolution: 1992–1998” technical manual on the HRS website.

HRS “Sampling Weights” technical manual on the HRS website.

Michael Hurd (2009?) “Invervivos Transfers and Bequests:A Little on Time Transfers” Slides presented for MacArthur Foundation Research Network conference.

Cindi Leacock, ed. (2006) “Getting Started with the Health and Retirement Study” version 1.0, technical manual on HRS website.

National Institute on Aging, “Growing Older in America:The Health & Retirement Study,” on HRS website.

Thank you