HEALTH AND RETIREMENT STUDY

Cross-Wave Geographic Information:
Respondent Census Division/Region and Mobility File

1992-2014

Data Description

Version 6.1, December 2016

Table of Contents

1. Overview

The Health and Retirement Study (HRS) is a national longitudinal study of the economic, health, marital, and family status, as well as public and private support systems, of older Americans. The HRS is a rich source of longitudinal, cross-sectional data for researchers and policymakers who study aging. Funding for the Health and Retirement Study is provided by the National Institute on Aging at NIH (U01 AG009740), with supplemental support from the Social Security Administration. The study is conducted by the Institute for Social Research (ISR) at the University of Michigan.

2. Setup

By downloading this freely provided data set, you agree to use its contents only for research and statistical purposes, making no effort to identify the respondents. You also agree inform HRS of any papers, publications, or presentations based on this data set. Please send a copy of such publications in PDF format via e-mail to *hrsquestions@umich.edu* with "Attn: Papers and Publications" in the subject line. If you wish, you may include a bibliographical reference.

As an alternative, you may transmit publications in paper format by postal mail:

```
Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3410
P.O. Box 1248
Ann Arbor, Michigan  48106-1248
```

2a. Distribution Set

The *Cross-Wave Geographic Information: Respondent Census Division/Region and Mobility* data set is packaged for distribution in a ZIP file, **HRSXRegion14.zip**, and contains region-level geographic information based on where the respondent was contacted in each interviewing year. Extract the data file, the program statement file that matches your analysis environment, the data description (this file), and the codebook file.

If you have problems when downloading this data set or in extracting its contents, please contact the HRS Help Desk. See Table 1 (below) for a description of the data set contents as well as a suggested subdirectory structure.

2b. Program Statements

Each data file comes with associated SPSS, SAS, or Stata program statements to read the data. Files containing SPSS statements are named with an .sps extension, those with SAS statements with an .sas extension, and those with Stata statements with .do and .dct extensions.

2b-1. Using the Files with SAS

To create a SAS system file for this data set, load the .sas program statement files into the SAS Program Editor and reference the appropriate .da data files. If the *.sas file is located in 'c:\xyrreg\sas" and the data file is located in 'c:\xyrreg\data", you can run the file as is. A SAS system file will be saved to directory 'c:\xyrreg\sas". If the files are not located in the specified directories, you will need to edit the *.sas file to reflect the proper path names prior to running the file.

2b-2. Using the Files with SPSS

To create an SPSS system file for this data set, load the .sps program statement files into the SPSS syntax editor window, reference the appropriate .da data files, and select the *Run>All* option. If the *.sps file is located in 'c:\xyrreg\spss" and the data file is located in 'c:\xyrreg\data", you can run the file as is. An SPSS system file (*.sav) will be saved to directory 'c:\xyrreg\spss". If the files are not located in the specified directories, you will need to edit the *.sps file to reflect the proper path names prior to running the file.

2b-3. Using the Files with Stata

To use Stata with this data set, three file types must be present for that data set: .dct, .do, and .da. Files with the suffix ".da" contain the raw data for Stata to read. Files with the suffix ".dct" are Stata dictionaries used by Stata to describe the data. Files with the suffix ".do" are short Stata programs ("do files") which you may use to read in the data. Load the .do file into Stata and then submit it. If the *.do and .dct files are located in 'c:\xyrreg\stata" and the data file is located in 'c:\xyrreg\data", you can run the .do file as is. If the files are not located in these directories, you must edit the *.do and *.dct files to reflect the proper path names before you run the files.

Table 1: Contents of the Distribution Package

Directory[1]	File	Type
`c:\xyrreg\data\`	`HRSXREGION14.da`	`Data files (ASCII text)`
`c:\xyrreg\docs\`	`HRSXREGION14.txt` `HRSXRegionDD_2014.pdf`	`Codebook file (ASCII text)` `Data Description (this document)`
`c:\xyrreg\sas\`	`HRSXREGION14.sas`	`SAS program statements`
`c:\xyrreg\spss\`	`HRSXREGION14.sps`	`SPSS program statements`
`c:\xyrreg\stata\`	`HRSXREGION14.dct/` `HRSXREGION14.do`	`Stata dictionary/"do" files`

2c. Non-Windows Environments

Non-Microsoft users should modify the default Windows file structure syntax to match that of their own operating system. The following examples should work for both Macintosh OS X and any Unix/Linux distribution. Open the SAS program file(s), SPSS syntax file(s) or the Stata do/dct files in an ASCII editor and make the changes indicated below.

SPSS in an OSX environment

In this example, we assume that the user has downloaded the region dataset and placed the files in a **Desktop** folder called **Region14** with the ASCII data file stored in subfolder **data** and the syntax file in subfolder **spss**. Then the commands in the syntax file would be modified to look like this:

```
FILE HANDLE xyrreg /name='Desktop/Region14/data/HRSXREGION14.da' LRECL=721.
DATA LIST FILE= xyrreg/
HHID 1-6(A)
[rest of syntax file goes here]
.
execute.
SAVE /outfile 'Desktop/Region14/spss/HRSXREGION14.sav'. Execute.
```

STATA in an OS X Environment

In the following example we assume that:
- The username is "user1"
- The zip file containing tracker information has been downloaded to the user's desktop from the HRS file download site
- The user has decompressed the zip file (use Stuffit for OS X) into a desktop folder named **Region14**
- The statistical package is stata

---

[1] When using HRS data products, you should feel free to create the directory structure that is most suitable for your needs. By using the suggested directory structure (or a Unix equivalent), you will not have to change the path name references in the data descriptor files. If you want to use a different structure, just change the directory references in the program files.

File `HRSXREGION14.do` should be modified as follows:

Change…
```
        infile using c:\xyrreg\stata\HRSXREGION14.dct
```
To…
```
        infile using /Users/user1/Desktop/Region14/stata/HRSXREGION14.dct
```

Change…
```
        save c:\xyrreg\stata\HRSXREGION14.dta
```
To…
```
        save /Users/user1/Desktop/Region14/stata/HRSXREGION14.dta
```

File `HRSXREGION14.dct` should be modified as follows

Change…
```
        dictionary using c:\xyrreg\data\HRSXREGION14.da {
```
To…
```
        dictionary using /Users/user1/Desktop/Region14/data/HRSXREGION14.da {
```

3. Data File Contents

The *Cross-Wave Geographic Information: Respondent Region and Mobility* data set replaces all previously released HRS region-level geographic information products. The new data set matches the current version of the tracker file (Tracker 2014)[2]. It has 38,183 records, which are uniquely identified by Household Identifier (HHID) and Person Number (PN).

This data set is released in conjunction with four other data sets:
- *Cross-Wave Geographic Information (Detail)* [restricted]
- *Cross-Wave Geographic Information (State)* [restricted]
- *Child ZIP Codes* [restricted]
- *Parent State Code* [restricted]

These files contain all geographic information (public and restricted) currently available for HRS respondents. See Table 2 for a comparison of the region, state, and detail file contents.

3a. Data Sources and Matching

In the earlier waves of the study (1992, 1993, 1994, 1995, 1996) geographical data were obtained from a variety of sources:
- HRS address/control files
- AHEAD address/control files
- The original HRS sample screen file
- Interview Content

Geographic information detail information was generated using these classification tools:
- the TIGER/Census Street Index
- the Federal Financial Institutions Examination Council Web site
- the Census Bureau Tract Street Locator
- CDUSA 9-Digit ZIP Code/Street Address listings

---

[2] The HRS tracker file is created to facilitate the use of HRS data within and across waves. It contains one record for every person who was ever eligible to be interviewed in any wave. The tracker file version used in preparation of this dataset (Tracker 2014, Early v1) covers all types of interviews (core, exit, and post-exit) for 1992 through the 2014 interviewing year.

From 1998 onward, inputs to the geographical data coding process were obtained from one source, the Survey Research Operations (SRO) field control system. A respondent address table was created prior to each wave and was updated by interviewers during the data collection phase of the study. The table contained street address, city, state and ZIP code fields which served as inputs to address lookup and matching software.

The current version of the cross-wave detail file (from which this file is derived) represents a complete rework of the coding process. The respondent information in the SRO control files reviewed for accuracy and consistency, the goal being to determine the actual interview location. Valid lot, street, city, state and postal code data were developed for each wave in which a respondent was interviewed. Once the cross-wave database update was complete, SAS PROC GEOCODE was used to determine longitude/latitude coordinates for each wave address. See *Cross-Wave Geographic Information (Detail) Data Description and Usage -- Appendix A*, for detailed information about the processes used to obtain geographic information in each wave.

3b. Content

Table 2: File Content Comparison

Variables	Description	Region	State	Detail Section
HHID PN STUDY FIRSTIW	ID & Tracker Variables	✓	✓	A
AIWTYPE – OIWTYPE	Interview Type	✓	✓	A
BORNUS	Born US?	✓	✓	A
REGLIV10	Region Live When in School	✓		
REGIONB	Census Division Where Born	✓		
REGIONyy	Region/Division of Residence	✓		
BEALE1993_yy BEALE2003_yy BEALE2013_yy	HRS-Beale Rural Urban codes by wave using 1993, 2003, and 2013 coding	✓		
MOVEwwww	Move distance from previous to current wave	✓		
STABORN	State in which R was born		✓	A
COUNTRYB92	Country Born (1992-2000)		✓	A
COUINTRYB02	Country Born (2002-2014)		✓	A
WHRLIV10	Where Live When in School		✓	A
STATEUSPSyy	State USPS code (from address file)		✓	
STFIPSyy	State FIPS Codes (2010 Census)		✓	
YEAR	Interview year (wave) designator			B
STATEUSPS	State USPS code (from address file)		✓	B
ZIPCODE	ZIP Code (from address file)			B
STATEFIPS10	State FIPS Codes (2010 Census)			B
CNTYFIPS10	County FIPS Codes (2010 Census)			B
TRACT10	State-County FIPS Codes (2010 Census)			B
LINKCEN2010	State FIPS + County FIPS + Tract (2010 Census)			B
LINKCEN2000	State FIPS + County FIPS + Tract (2000 Census)			B
LINKCEN1990	State FIPS + County FIPS + Tract (1990 Census)			B
COUNTYNAME10	County Name (2010 Census)			B
URBRUR2013 URBRUR2003 URBRUR1992	Beale Rural Urban codes for2013, 2003, and 1993			B
RESCODE	PROC GEOCODE match results (_NOTES_)			B
VERSION	Latest Version	✓	✓	A

3b-1. Variables from the Tracker File

The variables STUDY, FIRSTIW and BORNUS are taken directly from the Tracker file. Also included are Interview Type variables (xIWTYPE), which indicate whether an interview was attempted for the wave in question, and what type of interview (core, exit, post-exit) was obtained.

3b-2. One-time Variables[3]

Background variables that contain geographic information are asked at the time of the respondent's first interview:

- RegionB: The state where the respondent was born, recoded to Census Division
- RegLiv10: The state where the respondent lived when 10 years old or in school, recoded to Census Division.

3b-3. Cross-Wave Region

The region-level (Census Division) information in this data set is based on the state where the respondent was interviewed.

3b-4. Rural-Urban Variables

This data set contains bracketed versions of the Beale Rural-Urban Continuum Codes (1993, 2003 and 2013 versions) that have been collapsed to protect respondent confidentiality.[4]

Table 3: Rural-Urban Continuum Codes

1993	*2003*	*2013*	*Description*
			Metro counties:
0, 1	1	1	Counties in metro areas of 1 million population or more
2	2	2	Counties in metro areas of 250,000 to 1 million population
3	3	3	Counties in metro areas of fewer than 250,000 population
			Non-metro counties:
4	4	4	Urban population of 20,000 or more, adjacent to a metro area
5	5	5	Urban population of 20,000 or more, not adjacent to a metro area
7	6	6	Urban population of 2,500 to 19,999, adjacent to a metro area
7	7	7	Urban population of 2,500 to 19,999, not adjacent to a metro area
8	8	8	Completely rural or less than 2,500 urban population, adjacent to a metro area
9	9	9	Completely rural or less than 2,500 urban population, not adjacent to a metro area

3b-5. Distance Calculations

Researchers frequently express interest in using HRS geographic identifiers to calculate how far respondents have moved between waves. For this public data set, between-wave respondent moves (in miles) were calculated from pairs of latitude/longitude values derived from street address centroids.

---

[3] In previous versions of this data set, region/division values for place of birth and location of schooling were coded inconsistently for a small subset of respondents. These inconsistencies have been corrected in this version.

[4] From http://www.ers.usda.gov/Briefing/Rurality/RuralUrbCon/ : "In earlier versions of the Rural-Urban Continuum Codes, metro areas with 1 million population or more were subdivided between central counties (Code 0) and fringe counties (Code 1). The Code 1 group has become much less meaningful in the last two censuses as more and more counties of large metro areas have been rated as central counties by OMB procedures. In 2000, only 1.6 percent of the population of large metro areas was in fringe counties. Therefore, this distinction has been dropped. Codes 0 and 1 have been combined, and the new code 1 represents all counties in metro areas of 1 million or more population." Additional information on the Rural-Urban Continuum Code is available in the following USDA report: *Rural-Urban Continuum Codes for Metro and Non-metro Counties, 1993*, Margaret A. Butler and Calvin L. Beale, Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture, Staff report No. 9425, September 1994.

Users are reminded that geographic information will be missing or incomplete for respondent addresses that are outside the United States.

4. If You Need to Know More

This document is intended to serve as a brief overview to provide guidelines for using the *Cross-Wave Geographic Information: Respondent Census Division/Region and Mobility File*. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

4a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access public data or to find out more about restricted data products and procedures, visit the [HRS Web site](#).

4b. Contact Information

If you need to contact us, you may do so by one of the methods listed below.

```
Internet: Help Desk at the HRS Web site (http://hrsonline.isr.umich.edu)

E-mail: hrsquestions@umich.edu

Postal Service:
      Health and Retirement Study
      The Institute for Social Research
      426 Thompson Street, 3050 ISR
      Ann Arbor, Michigan  48104
```