HEALTH AND RETIREMENT STUDY

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

1992-2020

Data Description

Version 9 (Early), May 2023

Table of Contents

1. Overview 2
2. Obtaining the Data 2
Table 1: Contents of the Distribution Package 2
3. Data File Contents 3
3a. Data Sources and Matching 3
3b. Content 4
Table 2: File Content Comparison 4
3b-1. Variables from the Tracker File 5
3b-2. One-time Variables obtained from the baseline interview 5
3b-3. Cross-Wave Region 5
3b-4. Rural-Urban Variables 5
Table 3: Rural-Urban Continuum Codes 6
3b-5. Distance Calculations 6
3b-6. Foreign Addresses 6
4. If You Need to Know More 6
4a. HRS Internet Site 6
4b. Contact Information 6
Appendix 7
Build Instructions: Program Statements 7
Using the Files with SAS 7
Using the Files with SPSS 7
Using the Files with Stata 7
Build Instructions: Non-Windows Environments 7
SPSS in an OSX environment 8
STATA in an OS X Environment 8

**** EARLY RELEASE ****

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. Obtaining the Data

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 to inform HRS of any papers, publications, or presentations based on this data set. You may send a bibliographical reference (including a URL link whenever possible) for each item to *hrsquestions@umich.edu* with "Attn: Papers and Publications" in the subject line. If possible, you should also include a PDF-formatted copy of the publication.

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 3450
P.O. Box 1248
Ann Arbor, Michigan  48106-1248
```

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.

Table 1: Contents of the Distribution Package

Folders	Sub-Folders	File	Type
ascii	data	HRSXREGION20.da	Data file (ASCII text)
	sas	HRSXREGION20.sas	SAS program statements
	spss	HRSXREGION20.sps	SPSS program statements
	stata	HRSXREGION20.di/do	Stata dictionary and "do" files
built	Sas	HRSXREGION20.sas7bdat	SAS system file
	Spss	HRSXREGION20.sav	SPSS system file
	Stata	HRSXREGION20.dta	Stata system file
docs		HRSXRegionDD_2020v9.pdf	Data Description (This document)
		HRSXREGION20.txt	Codebook file (ASCII text)

3. Data File Contents

The *Cross-Wave Respondent Region/Division 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[1]. It has 43,559 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

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 that began with the 2014 wave. Respondent information in the SRO control files was 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*, for detailed information about the processes used to obtain geographic information in each wave.

[1] 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 2020) covers all types of interviews (core, exit, and post-exit) for 1992 through the 2020 interviewing year.

3b. Content

Table 2: File Content Comparison

Variables	***Description***	***Region***	***State***	***Detail (Section)***
HHID	Household Identifier	✓	✓	A, B
PN	Person Number	✓	✓	A, B
YEAR	Wave year numeric			B
WAVE	Wave year alphabetic			B
STUDY	Study Membership	✓	✓	A
FIRSTIW	Baseline Wave	✓	✓	A
AIWTYPE – PIWTYPE	Interview Type	✓	✓	
IWTYPE	Interview Type (by wave)			B
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	✓		
RESCODEyy	Centroid match information	✓		
StaBorn	State Born		✓	A
CountryB92	Country Born (1992-2000 code frame)		✓	A
CountryB02	Country Born (2002-2018 code frame)		✓	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
STCTYFIPS10	State FIPS + County FIPS (2010 Census)			B
STATEFIPS10	State FIPS Code (2010 Census)			B
CTYFIPS10	County FIPS Code (2010 Census)			B
TRACT10	Census Tract (2010 Census)			B
STCTYFIPS20	State FIPS + County FIPS (2020 Census)			B
STATEFIPS20	State FIPS Code (2020 Census)			B
CTYFIPS20	County FIPS Code (2020 Census)			B
TRACT20	Census Tract (2020 Census)			B
LINKCEN2020	State FIPS + County FIPS + Tract (2020 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
MOVE_TRACT	R Move Since Last Wave - Tract			B
MOVE_COUNTY	R Move Since Last Wave - County			B
MOVE_CITY	R Move Since Last Wave - City			B
MOVE_STATE	R Move Since Last Wave - State			B
MOVE_STREET	R Move Since Last Wave - Street			B
MOVE_ZIP	R Move Since Last Wave - Zip			B

Variables	*Description*	*Region*	*State*	*Detail (Section)*
UrbRur2013 UrbRur2003 UrbRur1992	Beale Rural Urban codes for2013, 2003, and 1993			B
HRRID	Hospital Referral Region			B
HRRLOCATION	HRR Location			B
HSAID	Hospital Service Area			B
HSALOCATION	HSA Location			B
HRRID2019	Hospital Referral Region 2019			B
HRRLOCATION 2019	HRR Location 2019			B
HSAID 2019	Hospital Service Area 2019			B
HSALOCATION 2019	HSA Location 2019			B
RUCA1990	Primary And Secondary RUCA Codes - 1990 Census			B
RUCAPRIME2000	Primary RUCA Code - 2000 Census			B
RUCASEC2000	Secondary RUCA Code - 2000 Census			B
RUCAPRIME2010	Primary RUCA Code - 2010 Census			B
RUCASEC2010	Secondary RUCA Code - 2010 Census			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 obtained from the baseline interview[2]

Background variables that contain masked versions of the geographic information 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.[3]

[2] 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.

[3] 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*

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, the SAS **geodist** function[4] was used to calculate between-wave respondent moves (in miles) from pairs of latitude/longitude values derived from street address centroids.

```
move1214 = geodist(lat12,lng12,lat14,lng14,"M");
```

3b-6. Foreign Addresses

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 *Respondent Census Region/Division 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 (https://hrs.isr.umich.edu/help)

E-mail: hrsquestions@umich.edu

Postal Service:
      Health and Retirement Study
      The Institute for Social Research
      426 Thompson Street, 3450 ISR
      Ann Arbor, Michigan  48104
```

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.

[4] SAS(R) 9.4 Functions and CALL Routines: Reference, Fourth Edition: GEODIST Function

Appendix

The *Cross-Wave Geographic Information: Respondent Census Division/Region and Mobility* data set is packaged for distribution in a ZIP file, **`HRSXRegion20v9.zip`**. Most users will wish to use the pre-built SAS, Stata, or SPSS versions of this dataset. If so, you may ignore the rest of this section.

Users who wish to build their own SAS, Stata, or SPSS file should extract the ASCII 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.

Build Instructions: 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.

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.

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.

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.

Build Instructions: 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 **Region20** 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/Region20/data/HRSXREGION20.da' LRECL=721.
DATA LIST FILE= xyrreg/
HHID 1-6(A)
[rest of syntax file goes here]
.
execute.
SAVE /outfile 'Desktop/Region20/spss/HRSXREGION20.sav'. Execute.
```

STATA in an OS X Environment

In the following example we assume that:

- The username is "user1"
- The zip file containing region 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 **Region20**
- The statistical package is stata

File `HRSXREGION20.do` should be modified as follows:

Change…

```
infile using c:\xyrreg\stata\HRSXREGION20.dct
```

To…

```
infile using /Users/user1/Desktop/Region20/stata/HRSXREGION20.dct
```

Change…

```
save c:\xyrreg\stata\HRSXREGION20.dta
```

To…

```
save /Users/user1/Desktop/Region20/stata/HRSXREGION20.dta
```

File `HRSXREGION16.dct` should be modified as follows

Change…

```
dictionary using c:\xyrreg\data\HRSXREGION20.da {
```

To…

```
dictionary using /Users/user1/Desktop/Region20/data/HRSXREGION20.da {
```