HEALTH AND RETIREMENT STUDY

Child ZIP Codes 2004-2020

Restricted Data

Data Description

Version 9.0 (Early), May 2023

To the Restricted Data Investigator: This restricted data set is intended for exclusive use by you and the persons specified in the *Agreement for Use of Restricted Data from the Health and Retirement Study* and/or the *Supplemental Agreement with Research Staff for Use of Restricted Data from the Health and Retirement Study*.

If there are any questions about this data set and its use, refer to the HRS Restricted Data Web Site (http://hrsonline.isr.umich.edu/rda) or contact the HRS Help Desk (hrsquestions@umich.edu).

This document may not be reproduced without the written consent of the staff of the Health and Retirement Study, The Institute for Social Research, The University of Michigan.

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. Obtaining the Data

2a. Access to Restricted Geographic Data

Although most HRS data sets are available to the public without restriction, certain HRS data sets contain sensitive respondent information and are only available under terms of a formal agreement negotiated between the researcher and HRS. Prospective users of HRS restricted geocode data have two access options:

- MiCDA Enclave Virtual Desktop Infrastructure (recommended)
- Traditional Licensing Agreement (deprecated)

For instructions on how to proceed, visit the HRS Restricted Data Web site or contact the HRS Restricted Data Applications Processing Team (*hrsrdaapplication@umich.edu*) by email.

2b. Restricted Data Agreement

This restricted data set is intended for exclusive use by you and the persons specified in the *Confidentiality Agreement for Use of Restricted Data from the Health and Retirement Study*. Traditional licensing agreements that include multiple users are also bound by the *Supplemental Agreement with Research Staff for Use of Restricted Data from the Health and Retirement Study*.

2c. Publications Based on Restricted Data

Your restricted data agreement requires that you inform HRS of any papers, publications, or presentations based on this restricted data set. You may send a bibliographical reference (including a URL link whenever possible) for each item to hrsrdaapplication@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
```

3. Content Overview

In Core interviews since 2004, respondents have been asked to provide location information (City, State, ZIP code) for children living more than 10 miles away. The format of these questions has changed over time:

3a. Question Sequence for 2004 through 2008

Prior to 2010, the respondent was asked to provide a ZIP Code for the child (E145). If the ZIP code was not given, the respondent was asked for city and state (E146/E147/E148). If the ZIP code was given, city and state were obtained from external sources.[1] If the ZIP code was not provided by the respondent, it was assigned from city and state information. Foreign addresses were dealt with separately in E149.

In the 2004 Core interview, the sequence was asked for each living child or child's surviving spouse/partner who was in contact and non-resident and who did not live within 10 miles of R. In the 2006 Core and 2008 Core interviews the same sequence was asked for new children or in situations where a child had moved since the previous interview.

3b. Question Sequence for 2010 through 2020

For 2010 onward, respondents were initially asked for each child's city and state (E146/E147/E148). If "other country" was assigned, country name was collected in E149. If a domestic city and state was provided for the child, the respondent was asked for the ZIP code (E145). The asking criteria were revised to include:

- All new children and children who have moved since the last interview of the Family R
- If the respondent has moved, all children who now live more than 10 miles from the respondent
- All children living more than 10 miles from the respondent for whom residence location information has never been obtained
- All children living more than 10 miles from the respondent whose residence location information was obtained more than four years prior to the current wave.

3c. Data Processing Procedures

2004-2008. If ZIP code was provided and city/state missing, the ZIP code value was used to extract city/state from online databases. If both ZIP code and city/state were provided by the respondent, city/state values were cross-checked against the city/state values in the ZIP code database, and inconsistencies were resolved on a case-by-case basis. In situations where city and state were provided but ZIP code was missing, the city/state combination was used to determine ZIP. In situations where a city/state combination had multiple ZIP codes, the first one on the list was assigned by the coder. Foreign addresses were processed separately; the city (if present) and country name were checked for accuracy and modified as necessary.

[1] For example, the **Zip-Codes.com Standard ZIP Code** database

2010-2020. If the respondent provided city, state and ZIP code, the combination was cross-checked to ensure consistency. If the respondent only provided city and state, ZIP code was assigned from the Zip-codes.com database, and in the case of problems, online (e.g., Google) lookups. In situations where a city/state combination had multiple ZIP codes, the first one on the list was assigned by the coder. Foreign addresses were processed separately; the city (if present) and country name were checked for accuracy and modified as necessary. [2]

3d. Dataset Construction Summary

This distribution set contains a cross-wave file constructed from the 2004 through 2020 wave-level files. See Table 1 for a list of the files provided in this distribution. Content of the file set created by the match process is summarized in Table 2, below. Build steps:

1. Input wave-level files keyed as follows by wave:

Wave	*Household Identifier*	*Current Wave Sub-Household ID*	*Other Person Number*
2020	*HHID*	*RSUBHH*	*OPN*
2018	*HHID*	*QSUBHH*	*OPN*
2016	*HHID*	*PSUBHH*	*OPN*
2014	*HHID*	*OSUBHH*	*OPN*
2012	*HHID*	*NSUBHH*	*OPN*
2010	*HHID*	*MSUBHH*	*OPN*
2008	*HHID*	*LSUBHH*	*OPN*
2006	*HHID*	*KSUBHH*	*OPN*
2004	*HHID*	*JSUBHH*	*OPN*

2. For each wave-level file:
    a. Rename sub-household IDs (xSUBHH) for each wave to SUBHH
    b. Create unique identifier: HHID + SUBHH + OPN + YEAR
    c. Update/harmonize other variables – City, State, Zipcode, Centroid, Match Status, Coding Procedures.
3. Concatenate by-wave records, then sort on new identifier to remove duplicates.
4. Create distribution file set.

Table 1: Contents of Distribution Package

Directory	***File***	***Type***
c:\czip\	ChildZip2004_2020.zip	Distribution file
c:\czip\xWave\docs\	ChildZipXwave.txt ChildZipCode2004_2020DD.pdf	Codebook(ASCII text) Data Description
c:\czip\xWave\sas\	Childzipxwave.sas7bdat	Cross-Wave SAS system file
c:\czip\xWave\spss\	Childzipxwave.sav	Cross-Wave SPSS system file
c:\czip\xWave\stata\	Childzipxwave.dta	Cross-Wave Stata system file

[2] In pre-August 2018 versions of the 2010-2014 files, the OPN variable contained the child index number (MX054_MC / NX054_MC / OX054_MC) from the Core interview. This was incorrect; therefore the 2010-2014 files were altered to include the correct OPN value. The child index number now appears in variable SEQNUM.

Table 2: Cross-Wave Child ZIP Code Processing Summary (2004-2020)

YEAR=2004

ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	373	0	0	0	0	0	0	373
Valid Zip	0	7441	13425	0	235	0	0	21101
State Only	0	0	0	1593	0	0	0	1593
DK/NA	0	0	0	0	0	0	1689	1689
Total	373	7441	13425	1593	235	0	1689	24756

YEAR=2006

ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	55	0	0	0	0	0	0	55
Valid Zip	0	2100	3380	0	673	0	0	6153
State Only	0	0	0	562	0	0	0	562
DK/NA	0	0	0	0	0	0	394	394
Total	55	2100	3380	562	673	0	394	7164

YEAR=2008

ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	151	0	0	0	0	0	0	151
Valid Zip	0	1712	2273	0	555	0	0	4540
State Only	0	0	0	420	0	0	0	420
DK/NA	0	0	0	0	0	0	299	299
Total	151	1712	2273	420	555	0	299	5410

YEAR=2010								
ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	594	0	0	0	0	0	0	594
Valid Zip	0	5605	13267	0	2262	0	0	21134
State Only	0	0	0	1222	0	0	0	1222
DK/NA	0	0	0	0	0	212	605	817
Total	594	5605	13267	1222	2262	212	605	23767

YEAR=2012								
ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	712	0	0	0	0	0	0	712
Valid Zip	0	6144	17141	0	23	0	0	23308
State Only	0	0	0	1258	0	0	0	1258
DK/NA	0	0	0	0	0	205	707	912
Total	712	6144	17141	1258	23	205	707	26190

YEAR=2014								
ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	825	0	0	0	0	0	0	825
Valid Zip	0	5390	13216	0	2977	0	0	21583
State Only	0	0	0	1256	0	0	0	1256
DK/NA	0	0	0	0	0	65	494	559
Total	825	5390	13216	1256	2977	65	494	24223

YEAR=2016								
ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	881	0	0	0	0	0	0	881
Valid Zip	0	6038	13251	0	1631	0	0	20920
State Only	0	0	0	1462	0	0	0	1462
DK/NA	0	0	0	0	0	732	0	732
Total	881	6038	13251	1462	1631	732	0	23995

YEAR=2018								
ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	838	0	0	0	0	0	0	838
Valid Zip	0	5780	10121	0	2344	0	0	18245
State Only	0	0	0	1304	0	0	0	1304
DK/NA	0	0	0	0	0	615	0	615
Total	838	5780	10121	1304	2344	615	0	21002

YEAR=2020								
ZIP CODE	CODING PROCEDURES FLAG							
	Foreign Country (0)	Zip/City/ State (1)	City / State (3)	State Only (4)	Other-resolved (5)	Other-Not resolved (6)	Incomplete/ DK/NA (9)	Total
Foreign	783	0	0	0	0	0	0	783
Valid Zip	0	5094	9550	0	2352	0	0	16996
State Only	0	0	0	1037	0	0	0	1037
DK/NA	0	0	0	0	0	810	0	810
Total	783	5094	9550	1037	2352	810	0	19626

4. If You Need to Know More

This document is intended to serve as a brief overview that provides guidelines for using this data product. 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

Appendix A: Installation

A1. Distribution Set

The *Child ZIP Code* data set is packaged for distribution in a .ZIP file, `ChildZip2004_2020.zip`. In order to keep the contents secure, the ZIP file has been encrypted using **WinZIP** 256 bit AES encryption. If you require a special file format or experience system problems, please contact the HRS Help Desk. If all files are decompressed, they will require approximately 165 MB of free space on your storage device.

A1-1. Windows Environment

Copy the ZIP file to the Windows folder where you plan to do your work. Use a third-party[3] file compression/decompression tool such as **WinZIP** or **7-zip** to extract the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table 1, above.

A1-2. UNIX/Linux Environment

Copy the ZIP file to the folder where you plan to do your work. Use the ZIP file decompression software installed on your system, (e.g. **7-zip, gunzip**) to decrypt and extract the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table 1, above.

A1-3. Macintosh OS X Environment (10.4.x and above)

Copy the ZIP file to the folder where you plan to do your work and use **Stuffit-Expander** to decrypt and extract the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table 1, above.

Extract the data file(s) matching your analysis environment, the data description (this file), and the codebook file(s).

Note: MiCDA Enclave Virtual Desktop Environment users are given access to pre-built SAS, Stata and SPSS versions of this dataset; therefore the installation information in this appendix does not apply to such users.

[3] The built-in Windows decompression utility will not process AES-256bit encrypted zip files; it halts with "an unexpected error is keeping you from copying the file".

Appendix B: Using the Child ZIP Code File for Analysis

B1. By-Wave Analysis:

The HRS household-member/child identifiers (OPN) were primarily designed to link the records with other records in a given wave; they were not optimized for merging records longitudinally across waves nor were they subjected to cross-wave consistency checks. Errors in identifiers have crept in across time. Given this, one option is to develop an analysis plan based on merging information from single-wave household-member/child records with single-wave respondent records and then merging the resulting respondent records longitudinally. Respondent records can reliably be merged longitudinally using Household Identifier (HHID) and Person Number (PN). We recommend this option if it accommodates your analytic needs. To extract Child ZIP Code information for an individual wave, filter on YEAR:

```
data mylib.hist2020_tmp; set mylib.childzipxwave; where  YEAR eq 2020; run;
```

B2. Cross-Wave Data Management using SAS.

The file that results from the process described in Section 3d, above can be modified to include variables and/or wave-years that are of analytic interest to you. Two examples of SAS procedures that create analysis files extending from 2020 backwards through time are provided below. If you are using Stata (see Section B3, below), you can specify equivalent statements.

Example 1: Create history of zip codes for OPNs in 2020 wave, transposing cross-wave ZIP data from long to wide.

```
* Transpose long to wide ;
proc transpose data=mylib.childzipxwave out=work.tmpOpnZip prefix=ZIP;
by hhid subhh opn;
id year; var ZIP; run;
proc sort data=work.tmpOpnZip; by hhid subhh opn; run;

* Select 2020 OPNs (n=19626) ;
data mylib.hist2020_tmp; set mylib.childzipxwave; where YEAR eq 2020; run;
proc sort data=mylib.hist2020_tmp out=mylib.hist2020; by hhid subhh opn; run;

* Merge to create cross-wave ZIP code file for 2020 OPNs   ;
data mylib.ziphist2004_2020;
RETAIN
HHID SUBHH OPN
ZIP2004 ZIP2006 ZIP2008
ZIP2010 ZIP2012 ZIP2014
ZIP2016 ZIP2018 ZIP2020 ;
Merge mylib.hist2020(in=a) mylib.tmpOpnZip(in=b) ;
by hhid subhh opn;
if a and b then output;
keep HHID SUBHH OPN ZIP2: ;
run;
```

Example 2: Create longitude/latitude historical file for 2020 OPNs. Since PROC TRANSPOSE has a one-variable limit, we get longitude, then latitude and merge the results before performing the 2020 linkage.

```
* Get longitude;
proc transpose data=mylib.childzipxwave out=work.wantA prefix=LONG;
by hhid subhh opn;
id year; var LONG;
run;

* get latitude ;
proc transpose data=mylib.childzipxwave out=work.wantB prefix=LAT;
by hhid subhh opn;
id year; var LAT;
run;

/* 2c. Put longitude and latitude together */
proc sort data=work.wantA; by HHID SUBHH OPN; run;
proc sort data=work.wantB; by HHID SUBHH OPN; run;

data work.longlat;
retain HHID SUBHH OPN
   LONG2004 LAT2004
   LONG2006 LAT2006
   LONG2008 LAT2008
   LONG2010 LAT2010
   LONG2012 LAT2012
   LONG2014 LAT2014
   LONG2016 LAT2016
   LONG2018 LAT2018
   LONG2020 LAT2020
   ;
Merge work.wantA(in=a) work.wantB(in=b) ;
by hhid subhh opn;
if a and b then output;
keep HHID SUBHH OPN LONG2: LAT2: ;
run;

* Select 2020 OPNs (n=19626);
data work.hist2020_tmp; set mylib.childzipxwave; where  YEAR eq 2020; run;
proc sort data=work.hist2020_tmp out=work.hist2020; by hhid subhh opn; run;

* Create history of zip centroids (Long/Lat) for the 2020 OPNs ;
data mylib.LongLat_2004_2020;
   retain HHID SUBHH OPN YEAR
   LONG2004 LAT2004 LONG2006 LAT2006
   LONG2008 LAT2008 LONG2010 LAT2010
   LONG2012 LAT2012 LONG2014 LAT2014
   LONG2016 LAT2016 LONG2018 LAT2018
   LONG2020 LAT2020 ;
merge work.hist2020_tmp(in=a) work.longlat(in=b) ;
by hhid subhh opn;
if a and b then output;
run;
```

B3. Cross-Wave Data Management using Stata.

```
* Example 1: Cross wave ZIP codes for each OPN
clear
* get child zipcode crosswave file
use D:\myfolder\CHILDZIPXWAVE.dta.

* select desired variable (ZIP) as well as ID variables
keep HHID SUBHH OPN YEAR ZIP

* convert long to wide
reshape wide ZIP, i(HHID SUBHH OPN) j(YEAR)
save D:\myfolder\widezip

* Example 2: Cross wave LONG/LAT for each OPN
clear
* get child zipcode crosswave file
use D:\myfolder\CHILDZIPXWAVE.dta.

* select desired variable (ZIP) as well as ID variables
keep HHID SUBHH OPN YEAR LONG LAT

* convert long to wide
reshape wide LONG LAT, i(HHID SUBHH OPN) j(YEAR)
save D:\myfolder\widelonglat

* Example 3: Subset 2020 OPNs
clear
* get child zipcode crosswave file 2020 records
use D:\myfolder\CHILDZIPXWAVE.dta if (YEAR==2020)
describe
* select the vars that we want to keep for 2020
keep HHID SUBHH OPN
describe
save D:\myfolder\wave2020

* Example 4: Merge datasets
merge 1:1 HHID SUBHH OPN using D:\myfolder\widezip.dta
save D:\myfolder\datamerged_data
```