

HEALTH AND RETIREMENT STUDY

Child ZIP Codes: 2004, 2006, 2008

Data Description and Usage

Version 3.0, November 2009
(Documentation revised August 2010)

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 at <http://hrsonline.isr.umich.edu/rda> or contact Michael A. Nolte at 734/936-1903 (voice) or manolte@isr.umich.edu (e-mail).

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 and Acknowledgments	2
2. Obtaining the Data	2
2a. Restricted Data Agreement	2
2b. Access to Restricted Data	2
2c. Publications Based on Restricted Data	3
3. Setup	3
3a. Distribution Set	3
3a-1. Windows Environment	4
3a-2. UNIX Environment.....	4
3b. Program Statements	5
3b-1. Using the Files with SAS.....	5
3b-2. Using the Files with SPSS	5
3b-3. Using the Files with Stata.....	5
4. Content Overview	5
5. If You Need to Know More	6
5a. HRS Internet Site	6
5b. Contact Information	7
Appendix: Cross-Wave Merges of Household-Member/Child Records.....	8
Sample SAS Code to Demonstrate Cross-Wave Merging.....	10

1. Overview and Acknowledgments

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.

The Health and Retirement Study began with two separate studies: the original HRS cohort, born between 1931 and 1941, first interviewed in 1992 (at 51-61 years of age); and the AHEAD (Asset and Health Dynamics among the Oldest Old) cohort, born before 1923 and first interviewed in 1993 (ages 70 and above). Both samples included spouses or partners, regardless of age.¹ In 1998, after two follow-ups of HRS (1994, 1996), and one of AHEAD (1995), the studies were merged and combined with new respondents born 1942-1947 (War Baby -- WB) and 1924-1930 (Children of the Depression Age -- CODA) to become a complete panel of the population over age 50. In 2004 a new sample cohort (Early Baby Boomer – EBB) of individuals born in 1948-53 (age 51-56 in 2004) was introduced. Future plans call for a new birth cohort to be enrolled every six years. All cohorts are drawn as nationally representative samples, with those based on the 1992 screen containing over-samples of African-Americans and Hispanic Americans.

2. Obtaining the Data

2a. Restricted Data Agreement

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 you have technical questions about the content and use of this data set, refer to the HRS Restricted Data Web site or contact Michael A. Nolte at 734.936.1903 (voice) or manolte@isr.umich.edu (email). All procedural questions should be directed to Cathy Liebowitz at 734.763.4180 (voice) or catlieb@isr.umich.edu (email).

2b. Access to Restricted 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. If you are a researcher who has found a [restricted data file](#) that is of interest, you should begin the application process by reviewing the

¹ The HRS sample came from a screening of 69,336 households conducted in 1992 and generated using a multi-stage, clustered area probability frame. The AHEAD sample was generated for what began as a separate study. Individuals born between 1914 and 1923, and about half of those born in 1913 or before, were identified through the same household screening used to identify the original HRS sample. The other half of those born in 1913 or before were identified using the Medicare enrollment files maintained by the Health Care Financing Administration (HCFA, since renamed the Centers for Medicare, Medicaid Services, or CMS).

application materials and procedures. If you decide to go forward, you should submit a preliminary application package containing a research proposal and data protection plan to HRS. If problems are found with your application, the HRS staff members will contact you to provide assistance in resolving them. Otherwise, once you have been notified by HRS that your application is complete, you may submit your application to your local Institutional Review Board and/or Contracting Authority. After IRB approval is obtained, you should submit the signed originals of the institutional approval documents to the HRS. At this point the application is complete and it is forwarded to the HRS Data Confidentiality Committee (DCC) for review and approval. If approval is granted by the DCC, the Principal Investigator of the HRS will sign the restricted data agreement, and you will be sent the restricted data product(s) by the HRS. Note: access codes for your restricted data product(s) are sent separately.

If your institutional or physical circumstances change, (e.g., new persons added to the project, hardware/software changes, or office moves) you should contact the HRS in order to modify the underlying agreement. A yearly re-certification of your restricted data agreement is required; if you wish to continue with analysis of the data after the expiration of the initial agreement, you will need to submit a renewal request. Finally, an HRS agent will periodically audit you in order to ensure that all conditions of the Restricted Data Agreement are being met.

2c. Publications Based on Restricted Data

As part of the data registration process, you have agreed to inform HRS of any papers, publications, or presentations based on this HRS restricted data set. Please send a copy of such publications, with a bibliographical reference if appropriate, to:

Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3050
P.O. Box 1248
Ann Arbor, Michigan 48106-1248

Alternately, you may contact us by e-mail at hqsquest@isr.umich.edu with "Attn: Papers and Publications" in the subject line.

3. Setup

3a. Distribution Set

The *Child ZIP Code* data set is packaged for distribution in a .ZIP file, `childzipCode.zip`, and contains city, state and ZIP Code information of children living more than 10 miles from their parent(s). This information is based on respondent reports for the 2004, 2006, and 2008 interviewing years (see Table 1 for a summary of package contents).

In order to keep the contents secure, the .ZIP file has been encrypted using PGP (Pretty Good Privacy). A PGP self-decrypting archive file is distributed for users working in a Microsoft Windows environment. Unix, Linux and Macintosh users receive a standard PGP file.² Extract the data file, the program statement file that matches your analysis environment, the data

² We assume that these users will have PGP installed on their systems. Users who require special file formats should contact the HRS Help Desk (hqsquest@isr.umich.edu).

description (this file), and the codebook file. If you have problems, please contact the [HRS Help Desk](#).

3a-1. Windows Environment

Run the self-decrypting archive file, `ChildZipCodeSda.exe`, from the command line or from Windows Explorer. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output from the self-extracting archive file will be the compressed file `ChildZipCode.zip`, which contains the files listed in Table 1. Use your file decompression software, (i.e. `Winzip`) to extract the files. At a minimum you will need to extract the data file, one set of program statement files, and the codebook file. Decompress the `.zip` file into the appropriate subdirectories (see Table 1 for a suggested structure). You should need no more than 10 MB of free space on your storage device to store all the files.

3a-2. UNIX Environment

Use the PGP software installed on your system to decrypt `ChildZipCode.pgp`. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output from the PGP file will be the compressed file `ChildZipCode.zip`, which contains the files listed in Table 1. Use your file decompression software, (i.e. `gunzip`) to extract the files. At a minimum you will need to extract the data file, one set of program statement files, and the codebook file. Decompress the `.zip` file into the appropriate subdirectories. You should need no more than 10 MB of free space on your storage device to store all the files.

Table 1: Contents of Distribution Package

Directory ³	File	Type
c:\czip\	ChildZip.zip	zip file
c:\czip\docs\	ChildZip2004.txt ChildZip2006.txt ChildZip2008.txt	Codebook files (ASCII text)
c:\czip\data\	ChildZip2004.da ChildZip2006.da ChildZip2008.da	Data files (ASCII text)
c:\czip\sas\	ChildZip2004.sas ChildZip2006.sas ChildZip2008.sas	SAS program statements
c:\czip\spss\	ChildZip2004.sps ChildZip2006.sps ChildZip2008.sps	SPSS program statements
c:\czip\stata\	ChildZip2004.dct/ChildZip2004.do ChildZip2006.dct/ChildZip2006.do ChildZip2008.dct/ChildZip2008.do	Stata dictionary and "do" files

³ While a particular setup is not required for using HRS files, we have traditionally suggested a directory structure for the Windows environment. By using this directory structure (or a Unix equivalent), you will not have to change the path name references in your data descriptor files. If you use a different structure, just change the directory references in the program files.

3b. 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. The files are keyed on Household Identifier, Other Person Number and Sub-Household Identifier. See the appendix for usage details.

3b-1. Using the Files with SAS

To create a SAS system file for a particular data set, two file types must be present for that data set -- .sas program statement files and .da data files. To create a SAS system file, load the *.sas file into the SAS Program Editor. If the *.sas file is located in "c:\czip\sas" and the data file is located in "c:\czip\data", you can run the file as is. A SAS system file (*.sd2, assuming that you are using SAS V6.12) will be saved to directory "c:\czip\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.

3b-2. Using the Files with SPSS

To create an SPSS system file for a particular data set, two file types must be present for that data set -- .sps program statement files and .da data files. To create an SPSS system file, open the *.sps file in SPSS as an SPSS Syntax File. If the *.sps file is located in "c:\czip\spss" and the data file is located in "c:\czip\data", you can run the file as is. An SPSS system file (*.sav) will be saved to directory "c:\czip\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.

3b-3. Using the Files with Stata

To use Stata with a particular data set, the following three file types must be present for that data set -- .dct files, .do files, and .da data files. 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:\czip\stata" and the data file is located in "c:\czip\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.

4. Content Overview

In the 2004 Core interview, respondents were asked to provide location information (City, State, ZIP Code) 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. Respondents were initially asked to provide the ZIP Code of the child's residence (E145). In situations where the respondent did not know the child's ZIP Code, city and state were asked (E146 and E147). If the respondent said that the child was living in a foreign country, the interviewer asked for the name of the country (E148/E149). 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.

If the respondent provided a ZIP Code for the child in E145, city and state were obtained from the Zip-Codes.com Standard ZIP Code database (<http://www.zip-codes.com>). If the ZIP Code

was not provided by the respondent, it was assigned from city and state information using the Zip-Codes.com database. In situations where a city/state combination had multiple ZIP Codes, the first one on the list was assigned by the coder.

Table 2: Child ZIP Code Processing Summary

ZIP Code Status	Coding Procedures Flag⁴						Total
	<i>Foreign Country</i> (0)	<i>Zip Code</i> (1)	<i>City & State</i> (3)	<i>State only</i> (4)	<i>Incomplete</i> (5)	<i>DK, NA</i> (9)	
2004 Core Interview							
<i>Foreign country (00000)</i>	373	0	0	0	0	0	373
<i>Valid ZIP Code (00612-99912)</i>	0	7441	13425	0	235	0	21101
<i>State only; no city or ZIP (99995)</i>	0	0	0	1593	0	0	1593
<i>DK; NA (99996)</i>	0	0	0	0	0	1689	1689
Total	373	7441	13425	1593	235	1689	24756
2006 Core Interview							
<i>Foreign country (00000)</i>	55	0	0	0	0	0	55
<i>Valid ZIP Code (00612-99912)</i>	0	2100	3380	0	673	0	6153
<i>State only; no city or ZIP (99995)</i>	0	0	0	562	0	0	562
<i>DK; NA (99996)</i>	0	0	0	0	0	394	394
Total	55	2100	3380	562	673	394	7164
2008 Core Interview							
<i>Foreign country (00000)</i>	151	0	0	0	0	0	151
<i>Valid ZIP Code (00612-99912)</i>	0	1712	2273	0	555	0	4540
<i>State only; no city or ZIP (99995)</i>	0	0	0	420	0	0	420
<i>DK; NA (99996)</i>	0	0	0	0	0	299	299
Total	151	1712	2273	420	555	299	5410

5. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines for using the *Child ZIP Code* 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.

5a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site at <http://hrsonline.isr.umich.edu/>.

Visit the HRS Restricted Data Web site at <http://hrsonline.isr.umich.edu/rda> if you want to find out more about restricted data products and procedures.

⁴ See codebook for details.

5b. Contact Information

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

Internet: [Help Desk](#) at our Web site

E-mail: hqsquest@isr.umich.edu

Postal Service:

Health and Retirement Study
The Institute for Social Research, Room 3050
The University of Michigan
P.O. Box 1248
Ann Arbor, MI 48106-1248

FAX: (734) 647-1186

Appendix: Cross-Wave Merges of Household-Member/Child Records⁵

The HRS household-member/child identifiers 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, you have a couple of options.

- Develop an analysis plan based on merging (summary) information from single-wave household-member/child records to 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.
- Merge household-member/child records longitudinally by using HHID, Sub-Household ID (xSUBHH), and Other Person Number (OPN), keeping in mind the following: When a household splits, a new xSUBHH value is generated for each respondent as two new households are formed. Child information collected for each household will be uniquely identified by combining OPN with the new xSUBHH value. This means that the same child may be referenced by two xSUBHH + OPN combinations⁶. It is also possible that OPN for a given child is not consistent from wave to wave. You should carefully consider how this limitation might affect your proposed analysis.

A process to create a file with ZIP Code information for 2002, 2004, 2006 and 2008 children is described below.

1. Link 2008 child records with 2006 child records using HHID, KSUBHH and OPN.
2. Link the 2008/2006 child records with 2004 child records using HHID, JSUBHH and OPN.
3. Link the 2008/2006/2004 child records with 2002 child records using HHID, HSUBHH and OPN.

At each step, if the earlier year data record is missing, fill in previous year's xSUBHH with the more recent year's xSUBHH, e.g., when merging 2008 records with 2006 records, `IF JSUBHH EQ ' ' THEN JSUBHH=KSUBHH`.

Wave	Current Wave Sub-Household ID	Previous Wave Sub-Household ID
2008	LSUBHH	KSUBHH
2006	KSUBHH	JSUBHH
2004	JSUBHH	HSUBHH
2002	HSUBHH	GSUBHH

⁵ We are indebted to Marita Servais for the explanation of cross-year child/other person merging techniques contained in this section.

⁶ The combination of xSUBHH + OPN is also referred to as Longitudinal Other Person Number (LOPN). For most types of merging you will need to use HHID, xSUBHH and OPN.

The file that results from this process can be used to subset records and create variables with information from earlier waves. SAS code to create a master file containing household-member/child records for all waves for all sub-samples is provided at the end of this document. You can modify it to include variables of analytic interest to you. If you are using SPSS or Stata, you would want to specify equivalent statements.

In 2004, child location information (City, State, ZIP Code) was asked for each living child or child's surviving spouse/partner who was in contact with the respondent, was non-resident, and who did not live within 10 miles of the respondent. In 2006 and 2008, location information was only asked for new children or in situations where a child had moved since the previous interview. So in order to categorize the 2008 children by ZIP Code you would need to merge information from 2004 and 2006, linking household-member/child records across time. See sample code at the end of this document for an example of how to carry out this process.

You should remember that duplicate records may occur because the household split in a subsequent year and provided more than one report for a child. In a household for a couple that split after the first wave of the study, a single child has two records in later waves - a "mom-says" record and a "dad-says" record. Thus in a given wave, one HHID + OPN can represent EITHER one person OR two different people.

The following illustration – couple divorces, one respondent remarries, both split-off households have new members – may help you visualize the situation. (Names are for illustration only.) In this example, after a split in 2006, a single OPN can link to two reports on one person (101), or separate reports for two different people (151). To avoid this situation, Joe and Amy's household receives the sub-household identifier KSUBHH=1. Joe's new stepchildren are assigned the sub-household identifier KSUBHH=1 and OPNs of 151 and 152. Carol's household receives the sub-household identifier KSUBHH=2. Carol's mother is also assigned a KSUBHH of 2 and an OPN of 151. Carol and Joe's daughter Susan keeps her original OPN of 101 and their respective reports on Susan are differentiated by the value of KSUBHH.

2004 Interview Year

Household records

HHID=090123 JSUBHH=0

Respondent records

HHID=090123 PN=010 JSUBHH=0 (Joe)

HHID=090123 PN=020 JSUBHH=0 (Carol)

Household member/child records

HHID=090123 JSUBHH=1 OPN=101 (daughter Susan)

2006 Interview Year

Household records

HHID=090123 KSUBHH=1

HHID=090123 KSUBHH=2

Respondent records

HHID=090123 PN=010 KSUBHH=1 (Joe)

HHID=090123 PN=011 KSUBHH=1 (Amy - new wife)

HHID=090123 PN=020 KSUBHH=2 (Carol)

Household member/child records

HHID=090123 KSUBHH=1 JSUBHH=0 OPN=101 (daughter Susan - Joe's report)

HHID=090123 KSUBHH=1 JSUBHH=0 OPN=151 (Joe's stepchild)
 HHID=090123 KSUBHH=1 JSUBHH=0 OPN=152 (Joe's stepchild)
 HHID=090123 KSUBHH=2 JSUBHH=0 OPN=101 (daughter Susan - Carol's report)
 HHID=090123 KSUBHH=2 JSUBHH=0 OPN=151 (Carol's mother)

Merged Master file

HHID=090123 OPN=101 KSUBHH=1 JSUBHH=0 DA04=1 DA06=1 (daughter Susan - Joe's report)
 HHID=090123 OPN=151 KSUBHH=1 JSUBHH=0 DA04=. DA06=1 (Joe's stepchild)
 HHID=090123 OPN=152 KSUBHH=1 JSUBHH=0 DA04=. DA06=1 (Joe's stepchild)
 HHID=090123 OPN=101 KSUBHH=2 JSUBHH=0 DA04=1 DA06=1 (daughter Susan - Carol's report)
 HHID=090123 OPN=151 KSUBHH=2 JSUBHH=0 DA04=. DA06=1 (Carol's mother)

Sample SAS Code to Demonstrate Cross-Wave Merging

```
*****;
*   Merging household-member/child records 2008 - 2004;
*****;
*   Create annual data files including HHID OPN cySUBHH pySUBHH,
    an assigned variable to indicate presence of record
    and variables of analytic interest, renamed if desired;
*
*****;

data da08;
  set cz2008.childzip2008;
  keep hhid opn lsubhh ksubhh da08 zip;
  da08=1;
  rename zip = zip08;
run;

data da06;
  set cz2006.childzip2006;
  keep hhid opn ksubhh jsubhh da06 zip;
  da06=1;
  rename zip = zip06;
run;

data da04;
  set cz2004.childzip2004;
  keep hhid opn jsubhh hsubhh da04 zip;
  da04=1;
  rename zip = zip04;
run;

*****;
*   Check distribution of analytic variables;

proc freq data=da08; table zip08 / missing nopercnt; run;
proc freq data=da06; table zip06 / missing nopercnt; run;
proc freq data=da04; table zip04 / missing nopercnt; run;

*****;
*   Combine biennial data files;
*****;
```

```

* Step 1;
proc sort data=da08; by hhid ksubhh opn; run;
proc sort data=da06; by hhid ksubhh opn; run;

data da0806;
  merge da08 da06;
  by hhid ksubhh opn;
  if jsubhh eq ' ' then jsubhh=ksubhh;
run;

* Step 2;
proc sort data=da0806; by hhid jsubhh opn; run;
proc sort data=da04; by hhid jsubhh opn; run;

data da0804;
  merge da0806 da04;
  by hhid jsubhh opn;
  if hsubhh eq ' ' then hsubhh=jsubhh;
run;

*****;
* depending on your analytic needs,
  create wave-specific historical files
  for merging with wave-specific household-member/child files;
*****;
* create historical file for 2008;
data hist2008;
  set da0804;
  * select records that will match a 2008 record;
  where da08 eq 1;

  * assign zipcode - most recent;
  lastzip="-----";
  if zip08 ne "" then lastzip=zip08;
  else if zip06 ne "" then lastzip=zip06;
  else if zip04 ne "" then lastzip=zip04;

  if lastzip eq "-----" then lastzip="99999";
run;

* eliminate duplicate records (from split household in subsequent year);
proc sort
  data=hist2008
  out=hist2008a nodupkey;
  by hhid lsubhh opn;
run;
*****;

```