HEALTH AND RETIREMENT STUDY

Cross-Wave Child ZIP Codes: 2004 through 2014

Data Description and Usage

Version 5.0, August 2018

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 specifies that you will inform HRS of any papers, publications, or presentations based on this restricted 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
```

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 2014

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

2010-2014. 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.

August 2018 Update. In previous 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 have been altered to include the correct OPN value. The child index number now appears in variable SEQNUM.

Table 1a: Child ZIP Code Processing Summary (2004-2008)

	Coding Procedures Flag[2]						
ZIP Code Status	*Foreign Country (0)*	*Zip Code (1)*	*City & State (3)*	*State only (4)*	*Incomplete (5)*	*DK, NA (9)*	*Total*
2004 Core Interview							
Foreign country (00000)	373	0	0	0	0	0	373
Valid ZIP Code (00612-99912)	0	7441	13425	0	235	0	21101
State only; no city or ZIP (99995)	0	0	0	1593	0	0	1593
DK; NA (99996)	0	0	0	0	0	1689	1689
Total	373	7441	13425	1593	235	1689	24756
2006 Core Interview							
Foreign country (00000)	55	0	0	0	0	0	55
Valid ZIP Code (00612-99912)	0	2100	3380	0	673	0	6153
State only; no city or ZIP (99995)	0	0	0	562	0	0	562
DK; NA (99996)	0	0	0	0	0	394	394
Total	55	2100	3380	562	673	394	7164
2008 Core Interview							
Foreign country (00000)	151	0	0	0	0	0	151
Valid ZIP Code (00612-99912)	0	1712	2273	0	555	0	4540
State only; no city or ZIP (99995)	0	0	0	420	0	0	420
DK; NA (99996)	0	0	0	0	0	299	299
Total	151	1712	2273	420	555	299	5410

---

[2] See codebooks for details.

Table 1b: Child ZIP Code Processing Summary (2010-2014)

	Coding Procedures Flag							
ZIP Code Status	*Foreign Country (0)*	*Zip Code (1)*	*City & State (3)*	*State only (4)*	*Incomplete, Resolved (5)*	*Not Resolved (6)*	*DK, NA (9)*	*Total*
2010 Core Interview								
Foreign country (00000)	594	0	0	0	0	0	0	594
Valid ZIP Code (00612-99912)	0	5605	13267	0	2262	0	0	21134
State only; no city or ZIP (99995)	0	0	0	1222	0	0	0	1222
DK; NA (99996)	0	0	0	0	0	212		212
No information provided (99999)	0	0	0	0	0	0	605	605
Total	594	5605	13267	1222	2262	212	605	23767
2012 Core Interview								
Foreign country (00000)	712	0	0	0	0	0	0	712
Valid ZIP Code (01002-99912)	0	6144	17141	0	23	0	0	23308
State only; no city or ZIP (99995)	0	0	0	1258	0	0	0	1258
DK; NA (99996)	0	0	0	0	0	205	0	205
99999	0	0	0	0	0	0	707	707
Total	712	6144	17141	1258	23	205	707	26190
2014 Core Interview								
Foreign country (00000)	831	0	0	0	0	0	0	831
Valid ZIP Code (00501-99912)	0	5394	13240	0	2980	0	0	21614
State only; no city or ZIP (99995)	0	0	0	1261	0	0	0	1261
DK; NA (99996)	0	0	0	0	0	66	0	66
No information provided (99999)	0	0	0	0	0	0	495	495
Total	831	5394	13240	1261	2980	66	495	24267

4. 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* files. 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 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.

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.

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
```

Appendix A: Installation (Traditional License Only)

1. Distribution Set

The *Child ZIP Code* data set is packaged for distribution in a .ZIP file, `ChildZip.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, 2008, 2010, 2012 and 2014 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 **WinZIP** 256 bit AES encryption. Extract the data file(s), the program statement file(s) matching your analysis environment, the data description (this file), and the codebook file. 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 40 MB of free space on your storage device.

1-1. 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 A1.

1-2. 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 A1.

1-3. 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 A1.

> **Note: MiCDA Enclave Virtual Desktop Environment users are given access to pre-built SAS, Stata and SPSS versions of this dataset; therefore the 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".

Table A1: Contents of Distribution Package

Directory[4]	File	Type
`c:\czip\`	`ChildZip.zip`	`zip file`

`c:\czip\docs\`	`ChildZip2004.txt`
	`ChildZip2006.txt`
	`ChildZip2008.txt`
	`ChildZip2010.txt`
	`ChildZip2012.txt`
	`ChildZip2014.txt`
		`Codebook files (ASCII text)`

`c:\czip\data\`	`ChildZip2004.da`
	`ChildZip2006.da`
	`ChildZip2008.da`
	`ChildZip2010.da`
	`ChildZip2012.da`
	`ChildZip2014.da`
		`Data files (ASCII text)`

`c:\czip\sas\`	`ChildZip2004.sas`
	`ChildZip2006.sas`
	`ChildZip2008.sas`
	`ChildZip2010.sas`
	`ChildZip2012.sas`
	`ChildZip2014.sas`
		`SAS program statements`

`c:\czip\spss\`	`ChildZip2004.sps`
	`ChildZip2006.sps`
	`ChildZip2008.sps`
	`ChildZip2010.sps`
	`ChildZip2012.sps`
	`ChildZip2014.sps`
		`SPSS program statements`

`c:\czip\stata\`	`ChildZip2004.dct/ChildZip2004.do`
	`ChildZip2006.dct/ChildZip2006.do`
	`ChildZip2008.dct/ChildZip2008.do`
	`ChildZip2010.dct/ChildZip2010.do`
	`ChildZip2012.dct/ChildZip2012.do`
	`ChildZip2014.dct/ChildZip2014.do`
		`Stata dictionary and "do" files`

2. Program Statements

ASCII data files are distributed with associated SPSS, SAS or Stata program statements to read the data and create a system files. 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.

2-1. Using the Files with SAS

To create a SAS system file for a particular 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:\czip\sas"

---

[4] 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.

and the data file is located in "c:\czip\data", you can run the file as is. A SAS system file (*.sas7bdat, assuming that you are using SAS V8.0 or higher) 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.

2-2. Using the Files with SPSS

To create an SPSS system file for a particular 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:\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.

2-3. Using the Files with Stata

To use Stata with a particular 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:\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.

3. 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 extracted the 2014 files from the *Child Zip Code* data set and placed the files in a **Desktop** folder called **czip14** 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 czip14 /name='Desktop/czip14/data/childzip2014.da' RECL=221.
DATA LIST FILE= czip14/
HHID 1-6(A)
[rest of syntax file goes here]
.
execute.
SAVE /outfile 'Desktop/czip14/spss/childzip2014.sav'.
Execute.
```

STATA in an OS X Environment

In the following example we assume that:
- The username is "user1"
- The encrypted zip file containing *Child Zip Code* data has been copied to the user's desktop from the CDROM sent by HRS.
- The user has decrypted /decompressed the zip file (use Stuffit for OS X) into a desktop folder named **czip14**
- The statistical package is stata

File `CHILDZIP2014.do` should be modified as follows:

Change…
```
    infile using c:\czip\stata\childzip2014.dct
```
To…
```
    infile using /Users/user1/Desktop/czip14/statachildzip2014.dct
```

Change…
```
    save c:\czip\stata\childzip2014.dta
```
To…
```
    save /Users/user1/Desktop/czip14/stata/childzip2014.dta
```

File `CHILDZIP2014.dct` should be modified as follows:

Change…
```
    dictionary using c:\czip\data\childzip2014.da {…}
```
To…
```
    dictionary using Users/user1/Desktop/czip14/data/childzip2014.da {…}
```

Appendix B: Cross-Wave Merges of Household-Member/Child Records[5]

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 household 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 combinations[6]. 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 2004 through 2014 children is described below.

1. Link 2014 child records with 2012 child records using HHID, NSUBHH and OPN.
2. Link the 2014/2012 child records with 2010 child records using HHID, MSUBHH and OPN.
3. Link the 2014/2012/2010 child records with 2008 child records using HHID, LSUBHH and OPN.
4. Link the 2014/2012/2010/2008 child records with 2006 child records using HHID, KSUBHH and OPN.
5. Link the 2014/2012/2010/2008/2006 child records with 2004 child records using HHID, JSUBHH 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 2006 records with 2004 records, IF HSUBHH EQ ' ' THEN HSUBHH=JSUBHH.

---

[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.

Wave	Current Wave Sub-Household ID	Previous Wave Sub-Household ID
2014	OSUBHH	NSUBHH
2012	NSUBHH	MSUBHH
2010	MSUBHH	LSUBHH
2008	LSUBHH	KSUBHH
2006	KSUBHH	JSUBHH
2004	JSUBHH	HSUBHH
2002	HSUBHH	GSUBHH

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 2014 - 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;
*
************************************************************************;
libname cz2014 'V:\SECURE\DropVars\childzip2014\sasdata';
libname cz2012 'V:\SECURE\DropVars\childzip2012\sasdata';
libname cz2010 'V:\SECURE\DropVars\childzip2010\sasdata';
libname cz2008 'V:\SECURE\DropVars\childzip2008\sasdata';
libname cz2006 'V:\SECURE\DropVars\childzip2006\sasdata';
libname cz2004 'V:\SECURE\DropVars\childzip2004\sasdata';

data da14;
     set cz2014.childzip2014;
     da14=1;
     keep hhid opn osubhh nsubhh dada14 zip;
     rename zip = zip14;
run;

data da12;
     set cz2012.childzip2012;
     da12=1;
     keep hhid opn nsubhh msubhh da12 zip;
     rename zip = zip12;
run;

data da10;
     set cz2010.childzip2010;
     da10=1;
     keep hhid opn msubhh lsubhh da10 zip;
     rename zip = zip10;
run;

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

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

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


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

proc freq data=da14; table zip14 / missing nopercent; run;
proc freq data=da12; table zip12 / missing nopercent; run;
proc freq data=da10; table zip10 / missing nopercent; run;
proc freq data=da08; table zip08 / missing nopercent; run;
proc freq data=da06; table zip06 / missing nopercent; run;
proc freq data=da04; table zip04 / missing nopercent; run;


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

*    Step 1;
proc sort data=da14; by hhid nsubhh opn; run;
proc sort data=da12; by hhid nsubhh opn; run;

data da1412;
     merge da14 da12;
     by hhid nsubhh opn;
     if nsubhh eq ' ' then nsubhh=osubhh;
run;


*    Step 2;
proc sort data=da1412; by hhid msubhh opn; run;
proc sort data=da10;   by hhid msubhh opn; run;

data da1410;
     merge da1412 da10;
     by hhid msubhh opn;
     if msubhh eq ' ' then msubhh=nsubhh;
run;


*    Step 3;
proc sort data=da1410; by hhid lsubhh opn; run;
proc sort data=da08; by hhid lsubhh opn; run;

data da1408;
     merge da1410 da08;
     by hhid lsubhh opn;
     if lsubhh eq ' ' then lsubhh=msubhh;
run;
```

```
*     Step 4;
proc sort data=da1408; by hhid ksubhh opn; run;
proc sort data=da06; by hhid ksubhh opn; run;

data da1406;
      merge da1408 da06;
      by hhid ksubhh opn;
      if ksubhh eq ' ' then ksubhh=lsubhh;
run;

*     Step 5;
proc sort data=da1406; by hhid jsubhh opn; run;
proc sort data=da04; by hhid jsubhh opn; run;

data da1404;
      merge da1406 da04;
      by hhid jsubhh opn;
      if jsubhh eq ' ' then jsubhh=ksubhh;
run;


***************************************************************************;
*     depending on your analytic needs,
      create wave-specific historical files
      for merging with wave-specific household-member/child files;

***************************************************************************;
*     create historical file for 2014;
data hist2014;
      set da1404;
      *     select records that will match a 2014 record;
      where da14 eq 1;

      *   assign zipcode - most recent;
      lastzip="-----";
      if zip14 ne "" then lastzip=zip14;
      else if zip12 ne "" then lastzip=zip12;
      else if zip10 ne "" then lastzip=zip10;
      else 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=hist2014
      out=hist2014a nodupkey;
      by hhid osubhh opn;
run;


***************************************************************************;
```