

# HEALTH AND RETIREMENT STUDY

## **Cross-Wave Geographic Information (Detail) Restricted Data**

1992-2014

Data Description and Usage

Version 6.1, December 2016

**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 ([hqsquestions@umich.edu](mailto:hqsquestions@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

<b>1. Overview .....</b>	<b>2</b>
<b>2. Obtaining the Data .....</b>	<b>2</b>
2a. Access to Restricted Geographic Data.....	2
2b. Restricted Data Agreement .....	2
2c. Publications Based on Restricted Data .....	2
<b>3. Data File Contents.....</b>	<b>2</b>
3a. Data Sources .....	3
Table 1: Section B Record Count by Wave and Study Membership .....	4
Table 2: File Content Comparison .....	5
3b. Section A Contents .....	5
3b1. Variables from Tracker File .....	5
3b2. One-time Variables.....	6
3c. Section B Contents .....	6
3c1. Census Tract Information .....	6
3c2. Other Geographic Identifiers.....	6
3c3. Rural-Urban Continuum Variables.....	6
Table 3: Rural-Urban Continuum Codes.....	6
3d. Match results.....	7
<b>4. Distance Calculations .....</b>	<b>7</b>
<b>5. If You Need to Know More .....</b>	<b>7</b>
5a. HRS Internet Site.....	7
5b. Contact Information .....	7
<b>Appendix A: Data Preparation and Matching Procedure .....</b>	<b>8</b>
1. Background .....	8
2. Cleaning Procedures .....	8
2a. Data Cleaning Levels.....	8
2b. Examples of Data Cleaning Corrections: .....	11
3. Special Matching Issues .....	12
3a. Multiple names for same object .....	12
3b. TIGER errors .....	12
3c. Missing Information.....	12
3d. Limitations due to TIGER update timing .....	12
3f. ZIP Code matching .....	12
<b>Appendix B: Installation (Traditional License).....</b>	<b>13</b>
B1. Distribution Set.....	13
B1-1. Windows Environment .....	13
B1-2. UNIX/Linux Environment .....	13
B1-3. Macintosh OS X Environment (10.4.x and above) .....	13
Table B-1: Contents of Distribution Package (Traditional License).....	13
B2. Program Statements.....	14
B2-1. Using the Files with SAS .....	14
B2-2. Using the Files with SPSS .....	14
B3-3. Using the Files with Stata.....	14
B3. Loading HRS Data Products in a non-Windows Environment.....	14
SPSS in an OSX environment .....	14
STATA in an OS X Environment .....	15

## 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 ([hrrdaapplications@umich.edu](mailto:hrrdaapplications@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. Send a bibliographical reference (including a URL link whenever possible) for each item to [hrrdaapplication@umich.edu](mailto:hrrdaapplication@umich.edu) with "Attn: Papers and Publications" in the subject line. You may 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 3410  
P.O. Box 1248  
Ann Arbor, Michigan 48106-1248

## 3. Data File Contents

The *Cross-Wave Geographic Information (Detail)* data set replaces all previously released HRS detail-level geographic information products. This release introduces a new file structure:

- **Section A** matches the current version of the tracker file (Tracker 2014)<sup>1</sup> (N=38,183) and contains respondent baseline information. Records are uniquely identified by Household Identifier (HHID) and Person Number (PN).
- **Section B** contains one record for each wave in which a respondent was interviewed (see table, below). Records are uniquely identified by Household Identifier (HHID), Person Number (PN) and Wave.

See Table 2 below, for additional information on the content of each level.

This data set is released in conjunction with four other data sets:

- *Cross-Wave Geographic Information (State)* [restricted]
- *Child ZIP Codes* [restricted]
- *Parent State Code* [restricted]
- *Cross-Wave Geographic Information: Respondent Region and Mobility* [public]

These files contain all geographic information (public and restricted) currently available for HRS respondents. See Table 3 for a comparison of the region, state, and detail file contents.

### 3a. Data Sources

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 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 represents a complete rework of the coding process. The 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 **Appendix A** for detailed information about the cleaning and matching process.

---

<sup>1</sup> 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.

Table 1: Section B Record Count by Wave and Study Membership

<b>Wave</b>	<b>Study Membership</b>						<b>Total</b>
	<i>HRS</i>	<i>AHEAD</i>	<i>CODA</i>	<i>War Baby</i>	<i>EBB</i>	<i>MBB</i>	
1992	12481	23	0	0	0	0	12504
1993	0	8351	0	0	0	0	8351
1994	12609	0	0	0	0	0	12609
1995	0	8379	0	0	0	0	8379
1996	12756	0	0	0	0	0	12756
1998	12461	7621	2383	2642	0	0	25107
2000	12189	6838	2389	2662	0	0	24078
2002	11777	5947	2328	2672	0	0	22724
2004	11225	4881	2165	2644	3438	0	24353
2006	10486	3807	1949	2596	3479	0	22317
2008	10035	3108	1807	2544	3468	0	20962
2010	9573	2460	1636	2499	4864	5213	26245
2012	8808	1821	1420	2405	4769	5295	24518
2014	8226	1394	1247	2322	4653	5299	23141
<b>Total</b>	132652	54630	17324	22986	24671	15807	268044

Table 2: File Content Comparison

<b>Variables</b>	<b>Description</b>	<b>Region</b>	<b>State</b>	<b>Detail Section</b>
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	✓		
MOVEwww	Move distance from previous to current wave	✓		
StaBorn	State Born		✓	A
CountryB92	Country Born (1992-2000 code frame)		✓	A
CountryB02	Country Born (2002-2014 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
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 for 2013, 2003, and 1993			B
RESCODE	PROC GEOCODE match results (_NOTES_)			B
VERSION	Latest Version	✓	✓	A

### 3b. Section A Contents

#### 3b1. Variables from Tracker File

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

### 3b2. One-time Variables<sup>2</sup>

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

- **Bornus**: Was the respondent born in the United States.
- **StaBorn**: The state where the respondent was born.
- **CountryB92**: The country where the respondent was born (1992-2000).
- **CountryB02**: The country where the respondent was born (2002 or thereafter).<sup>3</sup>
- **WhrLiv10**: The state or country where the respondent lived when 10 years old or in school.

### 3c. Section B Contents

#### 3c1. Census Tract Information

Tract-level information (State FIPS, County FIPS, Tract) is assigned by PROC GEOCODE and is based on 2010 Census tract boundaries. Researchers who wish to merge area data based on Census 2000 or Census 1990 with HRS geographic data should use LINKCEN2000 or LINKCEN1990.

#### 3c2. Other Geographic Identifiers

As noted in Table 2, this file contains additional geographic identifiers:

- State U.S. Postal Service Code (from address file)
- ZIP Code (from address file)
- County Name (linked on 2010 State+County FIPS)

#### 3c3. Rural-Urban Continuum Variables

Beale Rural-Urban Continuum Codes<sup>4</sup> based on 1993, 2003 and 2013 versions are provided for each wave.

Table 3: Rural-Urban Continuum Codes

1993	2003	2013	Description
			<i>Metro counties:</i>
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
			<i>Non-metro counties:</i>
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

<sup>2</sup> In previous versions of this data set, state values for STABORN and WHRLIV10 contained values for certain respondents that were coded in an inconsistent fashion. In the current version, the data and documentation have been made consistent.

<sup>3</sup> The code frame for CountryB02 is a super-set of the CountryB92 code frame.

<sup>4</sup> Complete information on Beale Rural Urban codes is available here: <http://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation.aspx>

### 3d. Match results

Users are reminded that geographic information will be missing or incomplete for respondent addresses that are outside the United States. In some cases, Census tract data fields are blank, although a record is provided. This indicates that address information was incomplete for the respondent or that the respondent was residing outside the United States during the interviewing period. See Appendix A for a discussion of the matching procedures used by HRS.

## 4. Distance Calculations

Researchers frequently express interest in using HRS geographic identifiers to calculate how far respondents have moved between waves. Between-wave respondent moves (in miles) can be calculated from pairs of latitude/longitude values radians by using (for example) the SAS **geodist** function:<sup>5</sup>

```
move1214 = round(geodist(lat12,lng12,lat14,lng14,"M"),.01);
```

Latitude and longitude can be derived from Census Tract or ZIP Code centroid information, available from a variety of Internet sources<sup>6</sup> as well as the SASHELP.ZIPCODE data set.<sup>7</sup>

## 5. If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines for using the *Cross-Wave Geographic Information (Detail)* data set. 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 public data and other relevant information, visit the [HRS Web site](#). If you want to find out more about restricted data products and procedures, visit the [HRS Restricted Data](#) pages.

### 5b. Contact Information

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

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

E-mail: [hqsquestions@umich.edu](mailto:hqsquestions@umich.edu)

Postal Service:

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

---

<sup>5</sup> [SAS\(R\) 9.4 Functions and CALL Routines: Reference, Fourth Edition: GEODIST Function](#)

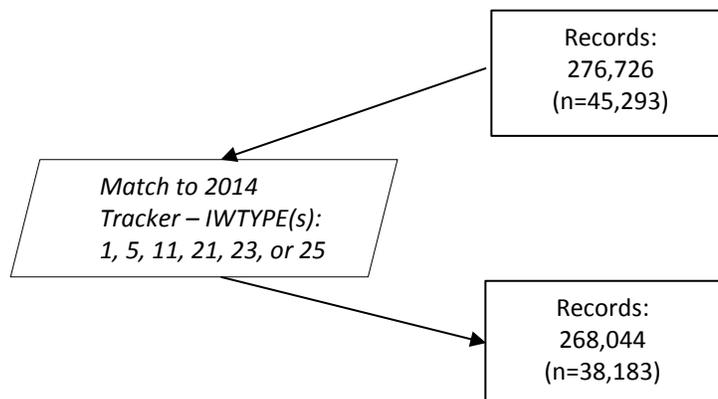
<sup>6</sup> See, for example, the [Missouri Census Data Center](#) Web site. Other sources include and the [U.S. Census Bureau](#) .

<sup>7</sup> See the [SAS Customer Support](#) site for [further details](#)

## Appendix A: Data Preparation and Matching Procedure

### 1. Background

Respondent addresses were compiled for study years 1992-2014, sourced from Field Control/Production database records. These records reflect respondent contact information at baseline entered by the interviewer. The respondent address information is confirmed with each subsequent interview wave contact. To build the cross-wave geographic detail file we concatenated wave-level records, producing a table of 276,726 records from 45,293 potential respondents. The 2014 Tracker file was used to select sample members for inclusion in the cross-wave file. At the wave level, valid respondents were defined as those with `xIWTYPE NE 99` (not in this wave).



At this point in the cleaning process, the addresses were raw and unstructured with varied styles (see below) that required comprehensive review and cleaning prior to using the SAS PROC GEOCODE<sup>8</sup> street-level matching procedure. PROC GEOCODE<sup>9</sup> street-level matching depends on consistent and accurate lot, street, city, and state information in order to achieve the highest possible assignment rate of longitude/latitude coordinates.

### 2. Cleaning Procedures

#### 2a. Data Cleaning Levels

- 1) High-level processing of the addresses in order to create inputs that would comply with USPS guidelines for proper structure. These modifications included:
  - a. Removing any punctuation including erroneous leading and superfluous spaces
  - b. Abbreviating the street prefix and/or street suffix e.g., "Street" to "ST", "ROAD" to "RD", "NORTH" TO "N", etc.
  - c. Correcting misspelled street and city names
  - d. Replacing business names i.e., nursing homes, long-term care, etc. facilities with the street address
  - e. Removing PO Box address
  - f. Removing secondary unit designators e.g., 3H, UNIT, APT, BLDG, SPC, LOT, REAR ENTRANCE, etc.
  - g. Ordinal indicators ("rd", "nd", "th") for numeric street names
  - h. Correcting invalid state USPS codes

<sup>8</sup> Massengill and Odom, *PROC GEOCODE: Now with Street-Level Geocoding* (SUGI 332-2010)

<http://support.sas.com/resources/papers/proceedings10/332-2010.pdf>

<sup>9</sup> [SAS Support: GEOCODE Procedure: Understanding Street Geocoding](#)

- i. Correcting invalid zip codes
- 2) In-depth cleaning was employed to achieve within respondent record homogeneity from wave to wave and achieve the greatest number of street-level matches. The SAS variables “\_NOTES\_” (*see table, below*) and “\_SCORE\_” were critical to identify records requiring more discernment. Post-geocoding match results of ZIP, CITY, and NONE underwent greater scrutiny and required the most cleaning. These modifications included:
- a. Correcting the street suffix to match the respondent’s preceding or following wave(s) addresses (example below)
  - b. Entering missing street suffix and/or prefix
  - c. Entering missing street directional
  - d. Entering missing ordinal indicators
  - e. Entering missing street addresses for a particular wave where IWTYPE = 1 and only when the preceding and following address matched (example below)
  - f. Entering missing zip codes (where a city has more than one zip code and only the respondent’s city and state were entered, the first zip code in ascending order was entered. For example: Peoria, AZ currently contains five zip codes: 85345, 85380, 85381, 85382, and 85383. If a respondent’s address only contained Peoria, AZ, then they were assigned 85345)
  - g. Correcting miscellaneous entry errors

SAS Variable _NOTES_ <sup>10</sup>		
This variable provides details about the quality of the address match by using token strings. Each token in _NOTES_ value has an associated score and the sum of scores make up the value of the _SCORE_ variable.		
TOKEN	VALUE	NOTE
AD	20	The street name matched.
CT	5	The city name matched.
DP	15	The street direction prefix matched.
DS	15	The street direction suffix matched.
ENDNM	0	The house number was outside the ranges of values in the lookup data set for the matching street. The geocoded coordinates for the nearest end of the street were used.
MCS	0	Multiple matches were found for the input street address and the street, city, and state in the street segment lookup data set.
MVP	0	The street geocoder detected missing values for the X or Y coordinates in the user-supplied lookup data set.
MZC	0	Multiple matches were found for the street address and ZIP code.
NM	10	The house number matched on the correct side of the street.
NMOS	5	The house number matched an address range in the lookup data set, but is on the opposite side of the street from the matched range.
NOADD	0	An invalid street address was input.
NOCTM	0	A match was found using the input address’ street and ZIP code. However, it was noted that the input city value was different from the city on the matched lookup data set observation.
NODAT	-10	The input address had no direction prefix but the matching street did have a direction prefix. For example, the input street name was "Main St." but the matching street was "N Main St."

<sup>10</sup> [SAS Support: GEOCODE Procedure: Understanding Street Geocoding: Street Geocoding Note Values](#)

NODPM	-15	The input address had a direction prefix but it either did not match the direction prefix of the matching street or the matching street had no direction prefix. For example, the input street name was "North Main St." but the matching street was "Main St."
NODSA	-10	The input address had no direction suffix but the matching street did have a direction suffix. For example, the input street name was "Johnson Ave" but the matching street was "Johnson Ave S."
NODSM	-15	The input address had a direction suffix but it either did not match the direction suffix of the matching street or the matching street had no direction prefix. For example, the input street name was "Johnson Ave South" but the matching street was "Johnson Ave."
NOLNM	0	The lookup data set contains missing values for the house numbers of the matching street. The geocoded coordinates for the center of the matching street were used.
NONM	0	The input address has no house number. The geocoded coordinates for the center of the matching street were used.
NOSTM	0	A match was found using the input address' street and ZIP code. However, it was noted that the input state value was different from the state on the matched lookup data set observation. For example, the input data set's address contained an incorrect state, "100 Main St., Cary, ND 27513". However, the matching lookup data set observation had the correct state value
NOTPA	-10	The input address had no street type prefix, but the matching address did have a street type prefix. For example, the input address was "110 Quebec." but the matching address was "110 Boulevard Quebec"
NOTPM	-20	The street type prefix of the input address was either not the same as the type prefix of the matching street or the matching street had no type prefix. For example, the input street name was "Boulevard Quebec" but the matching street name was "Avenue Quebec".
NOTSA	-10	The input address had no street type suffix, but the matching address did have a street type suffix. For example, the input address was "110 Main." but the matching address was "110 Main St."
NOTSM	-20	The street type suffix of the input address was either not the same as the type suffix of the matching street or the matching street had no suffix. For example, the input street name was "Park St." but the matching street name was "Park Ave."
NS	0	The input address had no state value
NSM	0	An initial match was found for the input city and country pair in the CITY lookup data set. However, the state variable value of the matching observation was not a match.
NOZC	0	No ZIP code was provided.
NOZCM	0	A match was found using the input address' street, city, and state. However, it was noted that the input ZIP code value was different from the ZIP on the matched lookup data set observation.
ST	5	The two-character state abbreviation matched.
TP	20	The street type prefix matched.
TS	20	The street type suffix matched.
ZC	15	The five-digit ZIP code or non-U.S. postal code matched.

2b. Examples of Data Cleaning Corrections:

2b-1. Change

<u>_SCORE_</u>	YEAR	STREET
<b>25</b>	1993	Main <b>CT</b>
<b>25</b>	1995	Main <b>CT</b>
65	1998	Main ST
65	2000	Main ST
65	2002	Main ST
65	2004	Main ST
65	2006	Main ST

"CT" changed to "ST"

2b-2. Added information

Street	City	State	Zip	Year	IWTYPE
8998 MAIN ST	Livonia	MI	48150	1994	1
8998 MAIN ST	Livonia	MI	48150	1996	1
	Livonia	MI	48150	1998	1
8998 MAIN ST	Livonia	MI	48150	2000	1

Becomes >>

Street	City	State	Zip	Year	IWTYPE
8998 MAIN ST	Livonia	MI	48150	1994	1
8998 MAIN ST	Livonia	MI	48150	1996	1
<b>*8998 MAIN ST</b>	Livonia	MI	48150	1998	1
8998 MAIN ST	Livonia	MI	48150	2000	1

\* Imputed street address

2b-3. Correction

Street	City	State	Zip	Year	IWTYPE
<b>1119</b> 2ND AVE	Detroit	MI	48201	1994	1
9111 2ND AVE	Detroit	MI	48201	1996	1
9111 2ND AVE	Detroit	MI	48201	1998	1
9111 2ND AVE	Detroit	MI	48201	2000	1

Corrected >>

Street	City	State	Zip	Year	IWTYPE
<b>9111</b> 2ND AVE	Detroit	MI	48201	1994	1
9111 2ND AVE	Detroit	MI	48201	1996	1
9111 2ND AVE	Detroit	MI	48201	1998	1
9111 2ND AVE	Detroit	MI	48201	2000	1

### 3. Special Matching Issues

#### *3a. Multiple names for same object*

There were numerous instances where a clean and valid address input to PROC GEOCODE resulted in a zip-level match instead of a longitude/latitude value. These were scrutinized to understand the reason for this limitation. Using the TIGERweb (<https://tigerweb.geo.census.gov/tigerweb/>) graphical representation of the lookup tables used by SAS PROC GEOCODE, it became apparent that a particular street/road might be assigned a second local or common use name. For instance, 123 Main St might also be known as 123 State Hwy 76. In similar fashion, a respondent might provide a street name, but the lookup data contain an alternative value such as highway route number. In these cases, the PROC GEOCODE information would be used in order to force a street-level match.

#### *3b. TIGER errors*

Also, there were times where the street name was misspelled on TIGERweb, therefore, the street name was changed in the data to force a street-level match. There were also instances where the street suffix had to be spelled out to achieve a street-level match.

#### *3c. Missing Information*

There were sparse occurrences of the street name simply not appearing on TIGERweb which resulted in a zip-level match only. Most of these were Courts or small road offshoots. Unfortunately, these could not be resolved and remain only a zip-level match. Also, addresses within private communities are often unlabeled in TIGERweb.

#### *3d. Limitations due to TIGER update timing*

At the date of this writing, it seems that newer (2012-2014) developments are not yet on the TIGER lookup tables and may result in only a zip-level match. Many of these developments appear unlabeled on the TIGERwebV8.0 satellite images, but not on the terrain view.

#### *3f. ZIP Code matching<sup>11</sup>*

If street level matching fails, PROC GEOCODE falls back to the use of ZIP code centroids. In the current file, Information matching based on ZIP code centroids may fail for several reasons:

- Missing data, coding and transcription errors
- Changes over time due to population shifts and changes to USPS internal procedures.
- Decommissioned zip codes

---

<sup>11</sup> Tracking historical changes to for a given ZIP code is uncertain since comprehensive cross-year references are not readily available.

## Appendix B: Installation (Traditional License)<sup>12</sup>

### B1. Distribution Set

The *Cross-Wave Geographic Information (Detail)* data set is packaged for distribution in a .ZIP file, `HRSXDetail.zip`. 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 40MB of free space on your storage device.

#### B1-1. Windows Environment

Copy the ZIP file to the Windows folder where you plan to do your work. Use a third-party<sup>13</sup> 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.

#### B1-2. 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.

#### B1-3. 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.

Table B-1: Contents of Distribution Package (Traditional License)

Directory <sup>14</sup>	File	Type
c:\xyrdet\data\	HRSXGEO14a_r.da HRSXGEO14b_r.da	Data files (ASCII text)
c:\xyrdet\docs\	HRSXGEO14.txt	Codebook file (ASCII text)
c:\xyrdet\sas\	HRSXGEO14a_r.sas HRSXGEO14b_r.sas	SAS program statements
c:\xyrdet\spss\	HRSXGEO14a_r.sps HRSXGEO14b_r.sps	SPSS program statements
c:\xyrdet\stata\	HRSXGEO14a_r.dct/HRSXGEO14a_r.do HRSXGEO14b_r.dct/HRSXGEO14b_r.do	Stata dictionary and "do" files

<sup>12</sup> 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

<sup>13</sup> 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".

<sup>14</sup> 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.

## B2. 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.

### *B2-1. 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:\xyrdet\sas" and the data file is located in "c:\xyrdet\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:\xyrdet\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.

### *B2-2. 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:\xyrdet\spss" and the data file is located in "c:\xyrdet\data", you can run the file as is. An SPSS system file (\*.sav) will be saved to directory "c:\xyrdet\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.

### *B2-3. 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:\xyrdet\stata" and the data file is located in "c:\xyrdet\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.

## B3. Loading HRS Data Products in a non-Windows Environment

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 .leveldo/.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 satate dataset and placed the files in a **Desktop** folder called **Geo14** with the ASCII data files stored in subfolder **data** and the syntax files in subfolder **spss**. Then the commands in the syntax file would be modified to look like this:

```
FILE HANDLE xyrdet /name='Desktop/Geo14/data/HRSXGE014A_R.da' RECL=53.
DATA LIST FILE= xyrdet/
HHID 1-6(A)
[rest of syntax file goes here]
.
execute.
SAVE /outfile 'Desktop/Geo14/spss/HRSXGE014_R.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 **Geo14**
- The statistical package is stata

File HRSXGEO14A\_R.do should be modified as follows:

*Change...*

```
infile using c:\xyrdet\stata\HRSXGEO14A_R.dct
```

*To...*

```
infile using /Users/user1/Desktop/Region14/stata/HRSXGEO14A_R.dct
```

*Change...*

```
save c:\xyrdet\stata\HRSXGEO14A_R.dta
```

*To...*

```
save /Users/user1/Desktop/Geo14/stata/HRSXGEO14A_R.dta
```

File HRSXGEO14A\_R.dct should be modified as follows

*Change...*

```
dictionary using c:\xyrdet\data\HRSXGEO14A_R.da {...}
```

*To...*

```
dictionary using Users/user1/Desktop/Geo14/data/HRSXGEO14A_R.da {...}
```