HEALTH AND RETIREMENT STUDY

Cross-Wave Geographic Information (Detail) All Cohorts, 1992-2022

Restricted Data

Data Description

Version 10, September 2025

To the Restricted Data Investigator: Use of this data product is limited to researcher(s) who have obtained authorization from the Health and Retirement Study and the University of Michigan.

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

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. Procedures

2a. Applying for Access

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 data may obtain access through the Michigan Center on the Demography of Aging (MiCDA) Virtual Desktop Infrastructure. For instructions on how to proceed, visit the HRS Restricted Data Web site. If you have questions, contact the HRS Restricted Data Applications Processing Team (hrsrdaapplication@umich.edu) by email.

Note: Use of each restricted data product is limited to researchers whose application for access to that product has been approved by the Health and Retirement Study and the University of Michigan.

2b. Audits

To help ensure the security of the restricted data and to prevent breaches of respondent confidentiality, HRS may conduct an audit. The audit process involves an unannounced site inspection of the work site(s) by an independent investigator, who will seek to verify compliance with the user's restricted data security plan. Violations will be the subject of further investigation and may result in withdrawal of data access, among other possible penalties.

2c. Publications Based on Restricted Data

All restricted data agreements require researcher to inform HRS of any papers, publications, or presentations based on restricted data sets. Researchers should 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. Whenever possible a PDF-formatted copy of the publication should be included. As an alternative, publications may be transmitted 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
```

2d. Modifications to Restricted Data Agreements

To make changes to an existing Restricted Data Agreement (RDA), or to make changes to a submitted RDA Application researchers must submit a Modify an Agreement form. The requirements, information, and documentation required for each type of modification are detailed on the Post-Application Activities Page. Typical modifications include annual renewal, research plan additions/updates, team member changes, and data protection plan modifications.

2e. Disclosure Limitation Review

A contractual obligation of researchers who qualify for access to restricted data from the Health and Retirement Study is to maintain respondent anonymity. Disclosure limitation review is the method by which HRS can prevent disclosure of confidential information, reduce the likelihood of respondent re-identification, provide useful data to researchers, and ensure the results of the review process are acceptable to both the researcher and the provider(s) of the restricted data. All materials proposed for import into and export from the Enclave are subject to disclosure review. For details on the review process, visit the Disclosure Limitation Review page.

3. Data File Contents

This version of the *Cross-Wave Geographic Information (Detail)* data set replaces all previously released HRS detail-level geographic information products. The 2010 and 2020 11-character tract linking variables in the previous version of this dataset contained missing data for certain tract edge cases. That problem has been corrected. In addition, Beale Rural-Urban Continuum codes for 2023 have been added. The detail data set employs a two-level file structure:

- **Section A** links to the current version of the tracker file (Tracker 2022)[1] and contains one baseline information record for each Tracker file respondent. All records are uniquely identified by Household Identifier (HHID) and Person Number (PN).
- **Section B** contains one record for each wave (1992 through 2022) in which a respondent was interviewed (see table, below) and address information was available. Records are uniquely identified by Household Identifier (HHID), Person Number (PN) and Wave.

See Tables 1 and 2 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:

[1] The HRS tracker file is created to facilitate the use of HRS data within and across waves. It contains one record for every person who was ever eligible to be interviewed in any wave. The tracker file version used in preparation of this dataset (Tracker 2022) covers all types of interviews (core, exit, and post-exit) for 1992 through the 2022 interviewing wave.

- 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.

In 2014, the geographic information coding process was reworked in conjunction with modifications to the SRO field control system The current version of the cross-wave detail file is based on respondent information in the SRO control files reviewed for accuracy and consistency in order to determine the actual respondent 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.

Note to users: Respondent addresses are carried forward to the next wave. Therefore they may appear for waves in which the interview did not occur. Tracker file variable xIWTYPE indicates whether someone was actually interviewed in a given wave, in which case the address would have been verified. **If users want to limit analyses to waves in which the R was interviewed they should retain waves of observations where xIWTYPE = 1.**
In summary, an address appearing on a given line does not indicate that an interview took place in that wave; it is simply the address that was on record at the time.

Table 1. Section A: Record Count by First Interview and Study Membership

Year	HRS	AHEAD	CODA	War Baby	EBB	MBB	LBB	EGENX	Total
1992	12521	131	0	0	0	0	0	0	12652
1993	0	8116	0	0	0	0	0	0	8116
1994	104	0	0	0	0	0	0	0	104
1995	0	107	0	0	0	0	0	0	107
1996	204	0	0	0	0	0	0	0	204
1998	195	59	2320	2529	0	0	0	0	5103
2000	121	19	43	87	0	0	0	0	270
2002	120	33	26	53	0	0	0	0	232
2004	95	19	8	24	3329	0	0	0	3475
2006	67	15	13	25	86	0	0	0	206
2008	49	7	7	23	57	0	0	0	143
2010	55	8	9	26	1386	4952	0	0	6436
2012	39	1	5	18	100	169	0	0	332
2014	20	2	2	17	42	95	0	0	178
2016	23	0	0	11	36	79	4417	0	4566
2018	9	0	1	3	18	21	46	0	98
2020	21	0	1	9	15	41	85	0	172
2022	91	2	12	143	285	409	389	1469	2800
9996	243	47	28	77	169	295	518	244	1621
Total	13977	8566	2475	3045	5523	6061	5455	1713	46815

Table 2. Section B: Record Count by Interview Year and Study Membership

Wave	HRS	AHEAD	CODA	War Baby	EBB	MBB	LBB	EGENX	Total
1992	12508	130	0	0	0	0	0	0	12638
1993	0	8231	0	0	0	0	0	0	8231
1994	12653	0	0	0	0	0	0	0	12653
1995	0	8382	0	0	0	0	0	0	8382
1996	12787	0	0	0	0	0	0	0	12787
1998	12496	7630	2390	2650	0	0	0	0	25166
2000	12233	6854	2396	2670	0	0	0	0	24153
2002	11827	5958	2337	2678	0	0	0	0	22800
2004	11286	4893	2172	2650	3446	0	0	0	24447
2006	10549	3819	1955	2602	3493	0	0	0	22418
2008	10095	3119	1814	2547	3480	0	0	0	21055
2010	9636	2470	1642	2502	4880	5228	0	0	26358
2012	8873	1830	1425	2409	4788	5302	0	0	24627
2014	8266	1402	1253	2323	4667	5311	0	0	23222
2016	7578	1014	1031	2233	4560	5300	4854	0	26570
2018	6698	686	806	2057	4350	5140	4928	0	24665
2020	6095	489	663	1963	4184	4997	4901	0	23292
2022	5269	303	488	1964	4262	5210	5200	1698	24394
Total	158849	57210	20372	31248	42110	36488	19883	1698	367858

Table 3. File Content Comparison[2]

Variables	***Description***	***Region***	***State***	***Detail (Section)***
HHID	Household Identifier	✓	✓	A, B
PN	Person Number	✓	✓	A, B
YEAR	Wave year numeric			B
WAVE	Wave year alphabetic			B
STUDY	Study Membership	✓	✓	A
FIRSTIW	Baseline Wave	✓	✓	A
AIWTYPE – PIWTYPE	Interview Type	✓	✓	
IWTYPE	Interview Type (by wave)			B
BornUS	Born US?	✓	✓	A
RegLiv10	Region Live When in School	✓		
RegionB	Census Division Where Born	✓		
REGIONyy	Region/Division of Residence	✓		
Beale1993_yy Beale2003_yy Beale2013_yy Beale2023_yy	HRS-Beale Rural Urban codes by wave using 1993, 2003, 2013, and 2033 coding	✓		
CHANGEwwww	Location change distance from previous to current wave	✓		
RESCODEyy	Centroid match information	✓		
StaBorn	State Born		✓	A
CountryB92	Country Born (1992-2000 code frame)		✓	A
CountryB02	Country Born (2002-2022 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
STCTYFIPS90	State FIPS + County FIPS (1990 Census)			B
STATEFIPS90	State FIPS Code (1990 Census)			B
CTYFIPS90	County FIPS Code (19900 Census)			B
TRACT90	Census Tract (2000 Census)			B
STCTYFIPS00	State FIPS + County FIPS (2000 Census)			B
STATEFIPS00	State FIPS Code (2000 Census)			B
CTYFIPS00	County FIPS Code (2000 Census)			B
TRACT00	Census Tract (2000 Census)			B
STCTYFIPS10	State FIPS + County FIPS (2010 Census)			B
STATEFIPS10	State FIPS Code (2010 Census)			B
CTYFIPS10	County FIPS Code (2010 Census)			B
TRACT10	Census Tract (2010 Census)			B
STCTYFIPS20	State FIPS + County FIPS (2020 Census)			B
STATEFIPS20	State FIPS Code (2020 Census)			B
CTYFIPS20	County FIPS Code (2020 Census)			B
TRACT20	Census Tract (2020 Census)			B

[2] See *Restricted/Redacted Geographic Variable Cross-Reference* (`RedactedGeocodeXref_v1.pdf`) for country and state code-frame information.

Variables	*Description*	*Region*	*State*	*Detail (Section)*
STCTYFIPS22	State FIPS + County FIPS (2022 Update)			B
STATEFIPS22	State FIPS Code (2022 Update)			B
CTYFIPS22	County FIPS Code (2022 Update)			B
TRACT22	Census Tract (2022 Update)			B
LINKCEN2022	State FIPS + County FIPS + Tract (2022 Update)			B
LINKCEN2020	State FIPS + County FIPS + Tract (2020 Census)			B
LINKCEN2010	State FIPS + County FIPS + Tract (2010 Census)			B
LINKCEN2000	State FIPS + County FIPS + Tract (2000 Census)			B
LINKCEN1990	State FIPS + County FIPS + Tract (1990 Census)			B
COUNTYNAME	County Name (PROC GEOCODE)			B
CHANGE_TRACT	Change Since Last Wave - Tract			B
CHANGE_COUNTY	Change Since Last Wave - County			B
CHANGE_CITY	Change Since Last Wave - City			B
CHANGE_STATE	Change Since Last Wave - State			B
CHANGE_STREET	Change Since Last Wave - Street			B
CHANGE_ZIP	Change Since Last Wave - Zip			B
UrbRur2023 UrbRur2013 UrbRur2003 UrbRur1992	Beale Rural Urban codes for 2023, 2013, 2003, and 1993			B
HRRID	Hospital Referral Region			B
HRRLOCATION	HRR Location			B
HSAID	Hospital Service Area			B
HSALOCATION	HSA Location			B
HRRID2019	Hospital Referral Region 2019			B
HRRLOCATION 2019	HRR Location 2019			B
HSAID 2019	Hospital Service Area 2019			B
HSALOCATION 2019	HSA Location 2019			B
RUCA1990	Primary And Secondary RUCA Codes - 1990 Census			B
RUCAPRIME2000	Primary RUCA Code - 2000 Census			B
RUCASEC2000	Secondary RUCA Code - 2000 Census			B
RUCAPRIME2010	Primary RUCA Code - 2010 Census			B
RUCASEC2010	Secondary RUCA Code - 2010 Census			B
RESCODE	PROC GEOCODE match results (_NOTES_)			B
VERSION	Latest Version	✓	✓	A

3b. Section A Contents

3b1. Variables from Tracker File

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

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).[3]
- 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 (LINK2010). Researchers who wish to merge area data based on other Census years with HRS geographic data should use LINKCEN2020, LINKCEN2000 or LINKCEN1990 as appropriate.[4]

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)
- Hospital Referral Region[5]
- Hospital Service Area
- Hospital Referral Region (2019)
- Hospital Service Area (2019)

[3] The code frame for CountryB02 is a super-set of the CountryB92 code frame.

[4] Users may note variations in 11-digit (State FIPS + County FIPS + Tract) with comparing 2010 with 2000 and/or 1990 tract links. This can result when the Census shapefiles are inconsistent with respect to a latitude/longitude centroid for an address that is on a tract boundary.

[5] Hospital Referral Region and Hospital Service Area codes are provided for users who wish to use Dartmouth Atlas of Health Care information in conjunction with HRS geographic data products.

3c3. Rural-Urban Continuum Variables

Beale Rural-Urban Continuum Codes[6] based on 1993, 2003, 2013 and 2033 versions are provided for each wave-level record.

Table 4a. Rural-Urban Continuum Codes (1993-2013)

1993	***2003***	***2013***	***Description***
			Metro counties:
0, 1	1	1	Counties in metro areas of 1 million population or more
2	2	2	Counties in metro areas of 250,000 to 1 million population
3	3	3	Counties in metro areas of fewer than 250,000 population
			Non-metro counties:
4	4	4	Urban population of 20,000 or more, adjacent to a metro area
5	5	5	Urban population of 20,000 or more, not adjacent to a metro area
7	6	6	Urban population of 2,500 to 19,999, adjacent to a metro area
7	7	7	Urban population of 2,500 to 19,999, not adjacent to a metro area
8	8	8	Completely rural or less than 2,500 urban population, adjacent to a metro area
9	9	9	Completely rural or less than 2,500 urban population, not adjacent to a metro area

Table 4b. Rural-Urban Continuum Codes (2023)

2023	***Description***
	Metro counties:
1	Metro - Counties in metro areas of 1 million population or more
2	Metro - Counties in metro areas of 250,000 to 1 million population
3	Metro - Counties in metro areas of fewer than 250,000 population
	Non-metro counties:
4	Nonmetro - Urban population of 20,000 or more, adjacent to a metro area
5	Nonmetro - Urban population of 20,000 or more, not adjacent to a metro area
6	Nonmetro - Urban population of 5,000 to 20,000, adjacent to a metro area
7	Nonmetro - Urban population of 5,000 to 20,000, not adjacent to a metro area
8	Nonmetro - Urban population of fewer than 5,000, adjacent to a metro area
9	Nonmetro - Urban population of fewer than 5,000, not adjacent to a metro area

3c4. Rural-Urban Commuting Area (RUCA) Codes

Rural-urban commuting area (RUCA) codes were developed by the USDA to classify U.S. census tracts using measures of population density, urbanization, and daily commuting. The most recent RUCA codes are based on data from the 2010 decennial census and the 2006-10 American Community Survey. The classification contains two levels. Whole numbers (1-10) delineate metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the primary (largest) commuting flows. These 10 codes are further subdivided based on secondary commuting flows, providing flexibility in combining levels to meet varying definitional needs and preferences.[7] Earlier versions of the RUCA codes (1990 and 2000) are also available and are included in this dataset.

[6] See https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/ for detailed information on Rural-Urban Continuum Codes.

[7] This paragraph is taken verbatim from the USDA web site. Visit https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ for further information.

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.

3e. Wave to Wave Comparison Variables

Between-wave changes in a respondent's address data (tract, county, city, state, street, ZIPcode) are recorded as CHANGE_TRACT, CHANGE_COUNTY, CHANGE_CITY, CHANGE_STATE, CHANGE_STREET, and CHANGE_ZIP.[8] Although any one of these variables can indicate a new address, users should not rely on an individual variable as the sole determinant of a respondent move. A combination of non-missing CHANGEx indicator variables should be used to make a final decision.

4. Distance Calculations

Researchers frequently express interest in using HRS geographic identifiers to calculate how far respondents have moved between waves. The SAS **geodist** function[9] allows the user to calculate between-wave respondent moves (in miles) can be calculated from pairs of latitude/longitude values.

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

Due to privacy concerns, HRS does not distribute latitude and longitude values specific to a respondent's address. However an approximation of respondent latitude and longitude can be derived from Census Tract or ZIP Code centroid information. Census tract centroid information is available from a variety of Internet sources,[10] For ZIP code centroids, a easily accessible source is the `SASHELP.ZIPCODE` data set.[11]

5. 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.

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 or to find out more about restricted data products and procedures, visit the HRS Web site.

[8] Note: In previous versions of this file, these variables were named MATCH_TRACT, MATCH_COUNTY, MATCH_CITY, MATCH_STATE, MATCH_STREET, and MATCH_ZIP. They have been renamed to align their content with usage strategies.

[9] SAS(R) 9.4 Functions and CALL Routines: Reference, Fifth Edition

[10] See, for example, the Missouri Census Data Center Web site as well as the U.S. Census Bureau .

[11] See ZIP Code 411: A Well-Kept SAS Secret, SUGI Paper 143-31 .

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 (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: Data Preparation and Matching Procedure

1. Background

Respondent addresses were compiled for study years 1992-2022, 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 2022 wave-level geographic detail file we concatenated control file wave-level records, producing a table of 367,938 records from 46,850 potential respondents (those with x`IWTYPE NE 99`). Please note that Section A of this data product is aligned with the 2022 Tracker File.

Control file addresses are raw and unstructured with varied styles (see below) that require comprehensive review and cleaning prior to using the SAS PROC GEOCODE[12] street-level matching procedure. It is important to note that PROC GEOCODE[13] 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. Record validation was conducted to remove null and invalid records which resulted in the final 1992-2022 detail file.

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
    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

[12] Massengill and Odom, *PROC GEOCODE: Now with Street-Level Geocoding* (SUGI 332-2010) https://support.sas.com/resources/papers/proceedings10/332-2010.pdf
[13] SAS Support: GEOCODE Procedure: Understanding Street Geocoding

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 zips 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_[14]		
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.
ADSDP	5	Address Direction Suffix matched lookup data Direction Prefix.
ADPDS	5	Address Direction Prefix matched lookup data Direction Suffix.
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. For example, the input data set's address contained an incorrect city, "100 Main St., Raleigh, NC 27513". However, the matching lookup data set observation had the correct city value, "100 Main St., Cary, NC 27513".
NODPA	-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."
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.

[14] SAS Support: GEOCODE Procedure: Understanding Street Geocoding: Street Geocoding Note Values

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, "100 Main St., Cary, NC 27513".
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. For example, the input data set's address contained a transposed ZIP code, "100 Main St., Cary, NC 25713". However, the matching lookup data set had the correct ZIP code value, "100 Main St., Cary, NC 27513".
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

SCORE	YEAR	STREET
25	1993	Main ***CT***
25	1995	Main ***CT***
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
****8998 MAIN ST***	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**
1119 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**
9111 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 TIGERweb, an online 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

According to the Census web site: "TIGERweb is updated twice a year. These updates are from the regular, semiannual snapshots of the MAF/TIGER system."[15] Given the complexity of the system it is possible that community-level changes are not immediately reflected in TIGER lookup tables. This may result in a PROC GEOCODE zip-level rather than street match.

3e. Post office box entries

Many zip-code level matches were a result of USPS postal box address entries. These types of mailing addresses should not have been entered in the respondent address fields.

3f. ZIP Code matching[16]

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

3g. Tract Information Assignment Sequence

PROC GEOCODE determines latitude/longitude centroid values for each matched address. The centroid value is then matched with Census 1990/2000/2010/2020 and 2022 shape files using PROC GINSIDE, the SAS point-in-polygon program. This process results in FIPS values for state, county, and tract.

[15] https://tigerweb.geo.census.gov/tigerwebmain/About_TIGERweb.pdf

[16] Tracking historical changes to for a given ZIP code is uncertain since comprehensive cross-year references are not readily available.