HEALTH AND RETIREMENT STUDY

2015 Life History Mail Survey (LHMS)

Restricted Data

Residential History Detail Level Data

Data Description and Usage

Version 2.0, May 2026

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).

Table of Contents

1. Overview

The *2015 Life History Mail Survey (LHMS) Version 2.0* contains questions about residential history, education history, and other important childhood and family events.

The LHMS is part of the Health and Retirement Study (HRS), which is funded under a cooperative agreement between the National Institute on Aging (NIA) and the Survey Research Center at the University of Michigan. The HRS is designed to study labor force, health, and family transitions of the U.S. population aged 51 and older, and the impact of those transitions on economic resources, claims on structured programs such as Social Security, Medicare, and Medicaid, and informal assistance and transfers to and from family members.

The National Institute on Aging (NIA) provided funding (R01 AG051142) for the 2015 LHMS, which was conducted by the Survey Research Center (SRC), at the Institute for Social Research (ISR), at the University of Michigan.

This dataset (lhms15_detail_reshist_v2) extends the LHMS by including ZIP code, city and state residential information. By receiving access to this dataset, you agree to use it for research and statistical purposes only and make no effort to identify respondents. In addition, you agree to send us a copy of publications you produce based on the data. See **Obtaining the Data** (at the end of this document) for additional details.

The questions in the 2015 LHMS are based on several similar studies around the world, including: the English Longitudinal Study of Aging (ELSA), England, The Survey of Health, Ageing and Retirement in Europe (SHARELIFE), and the China Health and Retirement Longitudinal Study (CHARLS).

The restricted data include the actual state indicators for the residential histories whereas the public release LHMS 2015 state data includes an enumeration of state residence.

2. Sampling Information

In December 2015, questionnaires were mailed to a subsample of HRS respondents (n= 11,256). The sample for the 2015 LHMS consists of all living HRS respondents who were not included in the 2015 Consumption and Activities Mail Survey (CAMS) and who completed their most recent HRS core interview in English (rather than Spanish). The field period for the 2015 LHMS was December 2015 through summer 2016.

3. File Structure

The *2015 HRS LHMS Geographic Detail Level* records are presented at the resident address-level. Street address table contain questions that were asked of all respondents about themselves (or were answered by a proxy for the respondent if the respondent was not able to fill out the questionnaire). The data file contains residential history records for respondents (including proxy responses) who returned a 2015 LHMS questionnaire. Of the 6,480 respondents who received a questionnaire, 5,860 (90.4%) provided at least one residential history response and are included in this file.

The file contains 38,830 records from 5,860 respondents, with each respondent contributing one or more address entries and a corresponding start and/or end year. On average, respondents recorded 6.6 residential history entries (min = 1, max = 25).

The data files are named LHMS15_DETAIL_RESHIST_V2, loosely translated as: Life History Mail Survey 2015, geographic detail-level data, version 2.

The primary identification variables (IDs) are:

- HHID HOUSEHOLD IDENTIFIER
- PN PERSON NUMBER
- ENTRY ROW NUMBER

4. LHMS Content

4a. Files Distributed (see Appendix A for variable listing)

Directory	File	Type
.\docs\	LHMS15_detail_reshist_v2.pdf	Data Description
.\sas\	LHMS15_detial_reshist_v2.sas7bdat	SAS system file
.\code\	LHMS_transpose_macro.sas	SAS syntax file
.\stata\	LHMS15_detail_reshist_v2.dta	Stata system file

4b. Data Quality

All missing answers are coded as "Answer not given/Missing", including missing due to skip patters and missing due to non-response. In cases where the respondent recorded ditto marks (“), and provided either a start or end year, the value of the last observation was carried forward. Records missing both start and end year were removed from the final dataset.

The variable for reported cities includes not only incorporated places but also a range of other geographic references, such as counties, townships, villages, neighborhoods, hamlets, census-designated places (CDPs), unincorporated communities, boroughs, military instillations and terms (e.g., “Base”), and broader descriptions like “in the country” and references to regions (e.g., “Western, MA”). These responses were retained as originally reported to preserve the language and intent of respondents, rather than imposing standardized geographic classifications that might obscure local or subjective meaning.

Although these records represent historical residences, ZIP code cleaning and validation were based on current ZIP code standards. Therefore, standardized ZIP values should be interpreted as current or modern postal codes associated with the address location, not necessarily the ZIP codes that existed at the time the respondent lived there.

4c-0. LH5 Series, Residential History – ZIP CODE Processing

In Version 2, missing zip codes were not imputed by HRS staff, as a result many records will be missing zip code values. A systematic approach was applied (see Appendix B for decision hierarchy):

ZIP code processing began with address-related fields, including street, city, state, original respondent ZIP, and any existing ZIP value. A ZIP lookup table was loaded to identify the number of ZIP codes associated with each U.S. city/state combination and, where applicable, the first ZIP associated with a single-ZIP city.

Before ZIP lookup and geocoding, city and state values were cleaned. This included correcting common misspellings where they could be identified in a state-specific context, normalizing township and borough suffixes, and standardizing obvious city-name variants (e.g., BKLYN, NY to Brooklyn, NY).

Respondent ZIP codes were then parsed and standardized. Placeholder values, nonsensical entries, too-short values, and missing values were identified. Candidate five-digit ZIP codes were created when possible, including adding a leading zero in states where five-digit zip codes begin with zero. Candidate ZIP codes were checked against the reported state using ZIP/state validation.

After preliminary ZIP code cleaning, city names were further normalized for geocoding. For example, New York City boroughs and neighborhoods were standardized to “NEW YORK” in appropriate cases. City values that duplicated the state name or state abbreviation, as well as other non-geographic city tokens, were treated as missing when appropriate.

The address data were then passed through geocoding using PROC GEOCODE in SAS. Geocoding output included match type, matched address components, matched ZIP, and coordinates. After geocoding, additional ZIP filling was performed. If a record had a missing or invalid pre-geocode ZIP, matched at the street level, had a usable house number and street name, and received a matched ZIP from geocoding, the ZIP was filled from the geocoder output and flagged accordingly.

For street-level geocode matches that still lacked a ZIP code, coordinates were exported to a Python GeoPandas process. Point-in-polygon matching against ZIP boundary files was used to assign ZIP codes where possible. These zip codes were then merged back into the dataset and flagged.

Finally, ZIP code information was consolidated into ZIP_FLAG. This flag preserves the pre-geocode classification when no later update occurred, records ZIP codes determined from SAS geocoding or Python spatial matching, applies a small number of manual edge-case fixes, allows limited fallback filling for single-ZIP cities, and identifies ZIP codes filled from ZIP-level geocode matches. A companion indicator, ZIP_GENERATED, identifies ZIP codes that were generated or filled during processing rather than directly reported by the respondent.

4c-1. LH5 Series, Residential History – Start and End Years

For variables START and END, respondents were instructed to write the start and end year for each of their residences before age 50, and to write their age or the decade if they could not remember the exact year. Some respondents recorded consecutive years on each line and the same residence i.e., city and state names. In these cases, the reported residence information was retained as given, but the repeated annual entries were collapsed into a single residence record and the start and end years were adjusted to reflect the full span covered by the consecutive entries (see 4d).

In a very small number of cases (24 records) the reported end year precedes the start year. Each of these records was individually verified against the respondent's original questionnaire. Where the discrepancy existed on the source document the values were left unchanged in the data to preserve the original response. No corrections were applied to these records

4c-2. LH5 Series, Residential History – Data Corrections

Missing state values were imputed using the following correction scheme:

1. A valid zip code was provided that could be matched to a state
2. State information was provided in other address components (e.g., listed in address field instead of state field)
3. Visual scan of questionnaire showed state information

State code 'ZZ' denotes a foreign country and/or broad responses such as "Pacific theatre WWII" and "Overseas." Records that matched at the ZIP code level but contained incomplete address information were carefully reviewed and corrected, when possible, to achieve better precision. Efforts were made to ensure that each entry represents a unique location.

4d. LH5 Series, Residential History – Record Aggregation

Questionable responses were checked against questionnaire scans, and discrepancies between state codes and ZIP codes were investigated and resolved. The data were modified by removing records lacking both a start and end year and incorporating newly found records. To eliminate redundancy, records with repeating information were collapsed. See tables A and B (note: entry 11 in A is collapsed and merged with entry 10 on B):

A

Entry	start	end	street	city	state	zip
10	1995	2005		Detroit	MI	·
11	2006	***2012***		Detroit	MI	·
12	2015			Wixom	MI	·

→

B

entry	start	end	street	city	state	zip
10	1995	***2012***		Detroit	MI	·
12	2015			Wixom	MI	·

As a result of collapsing or merging duplicate consecutive records, ENTRY values are not always consecutive within respondent records. Gaps in ENTRY values do not necessarily indicate missing residences or moves; they may reflect records that were collapsed during processing or lines that were skipped or otherwise used by the respondent on the questionnaire. The ENTRY variable identifies only the original questionnaire line on which the information was recorded. Researchers should not use ENTRY to count residences, residential episodes, or moves.

There are 149 records missing state codes pertain to non-geographically specific military service responses, such as “US Army” and “Navy,” college/dorm, and other relevant answers.

5. Merge Restricted to Public Data

To facilitate merging the residential history file with the survey data, a SAS macro has been developed that converts the residential history file from long format to wide format producing one row per respondent with address variables suffixed by entry number (e.g., city_1, zipcode_1, city_2, zipcode_2). Once in wide format the file can be merged directly in the restricted data enclave with the LHMS public data using the household identifier (HHID) and person number (PN) without additional reshaping.

The macro is available as a supplemental file (lhms_transpose_macro.sas) and requires SAS 9.4 or later. No SAS programming experience is necessary to run it. A typical conversion requires three lines of code (see Appendix D for macro usage).

The residential history datasets are sorted lexically by HHID, PN, and ENTRY using SAS linguistic collation with numeric collation enabled. This ensures that numeric values embedded in character variables sort in natural numeric order rather than character order. For example, ENTRY = "2" sorts before ENTRY = "10" rather than after it as would occur under standard character sorting.

The variable PN refers to the respondent to whom the questionnaire was mailed. Please note that identification variables in the *HRS 2015 LHMS Geographic Detail Level* are stored in character format. For further information about HRS identification variables and merging data, see the Tracker Data Description and the Data Descriptions for each HRS Core wave.

6. Linking Respondents across Time

Respondent records in the HRS 2015 LHMS Geographic Detail Level V2 when transposed can be linked to respondent records from prior waves by HHID and PN. The core sub-household identifiers can be used to link household data with the cross-sectional respondent level data.

7. Obtaining the Data

7a. Access to Restricted Data

Although most HRS data sets are available to the public without restriction, certain HRS data sets contain sensitive respondent information and are only available under terms of a formal agreement negotiated between the researcher and HRS. 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.

7b. Restricted Data Agreement

This restricted data set is intended for exclusive use by you and the persons (if any) specified in your restricted data agreement.

7c. 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. You may send a bibliographical reference (including a

URL link whenever possible) for each item to hrsrdaapplication@umich.edu with "Attn: Papers and Publications" in the subject line. If possible, you should also include a PDF-formatted copy of the publication.

As an alternative, you may transmit publications in paper format by postal mail:

```
Health and Retirement Study
Attn: Papers and Publications
The Institute for Social Research, Room 3450
P.O. Box 1248
Ann Arbor, Michigan  48106-1248
```

8. 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.

8a. 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.

8b. Contact Information

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

```
Internet: Help Desk at the HRS Web site (https://hrs.isr.umich.edu/help)

E-mail: hrsquestions@umich.edu

Postal Service:
      Health and Retirement Study
      The Institute for Social Research
      426 Thompson St, 3450 ISR
      Ann Arbor, Michigan  48104
```

Appendix A – Variable Descriptions

#	Variable	Description
1	HHID	HRS – Household Identifier
2	PN	HRS – Three-digit person number within the household. Combined with HHID uniquely identifies a respondent
3	ENTRY	indicates the questionnaire line number on which the respondent recorded the residence. It is not a sequential counter of residences or moves and has no substantive meaning beyond identifying the original response line. Records are sorted lexically with numeric collation so ENTRY = "2" sorts before ENTRY = "10"
4	_MATCHED_	Match level returned by SAS PROC GEOCODE indicating the precision of the geocoded location. Common values include Street (highest precision), ZIP, City, and None (no match). See SAS PROC GEOCODE documentation for full value definitions
5	_SCORE_	PROC GEOCODE match score indicating the relative quality of the address match; higher scores represent better matches.
6	_NOTES_	Supplemental notes returned by SAS PROC GEOCODE describing match details or reasons a record could not be geocoded. (see Appendix B for details)
7	START	Year the respondent recorded as the start year for an address
8	END	Year the respondent stopped residing at this address. 24 records have END before START and were verified against the original questionnaire
9	CITY	Respondent-reported city name as cleaned and standardized during data processing. Borough and neighborhood names within New York City were standardized to NEW YORK. Non-geographic entries and entries matching state names were removed
10	STATE	Two-character USPS state abbreviation based on respondent provided information. Foreign and military addresses are coded ZZ
11	ZIPCODE	Five-digit ZIP code assigned through a multi-stage process. Sources include respondent self-report, SAS PROC GEOCODE street-level match, Python GeoPandas boundary lookup, and single-ZIP city lookup. See ZIP_FLAG for the source of each record's ZIP code. Missing when no ZIP could be reliably assigned
12	TRACT	Census 2020 six-digit tract code for the geocoded address. Available only for records that achieved a street-level geocode match.
13	COUNTYFP	Three-digit Census county FIPS code for the geocoded address
14	STATEFIPS	Two-character state FIPS code corresponding to the respondent-reported state
15	FOREIGN	Binary indicator. 1 = address is foreign or military (STATE = ZZ). 0 = domestic address. Records with FOREIGN = 1 have no ZIP, tract, or county assignment
16	REGION22	Numeric Census region-division code based on the 2022 Census regional classification scheme. Derived from the respondent-reported state
17	REGIONDIV_LABEL	Descriptive label for the Census region and division corresponding to REGION22. For example, East North Central or Pacific
18	ZIP_FLAG	Documents the method used to assign the ZIP code. Allows researchers to filter by ZIP source or exclude generated ZIP codes. See ZIP Flag documentation for full value descriptions
19	ADDR_SPECIAL_TYPE	Identifies addresses that are not standard street addresses. Values include Military (military or veteran address), Homeless (homeless or transitional housing), College (campus or dormitory address), and Orphanage/foster care. Blank for standard residential street addresses
20	ZIP_GENERATED	Binary indicator. 1 = ZIP code was assigned by an automated method (geocoder, city lookup, or Python boundary fill) rather than reported by the respondent or missing. 0 = ZIP code was provided directly by the respondent and passed validation.

Appendix B – SAS _NOTES_ Tokens[1]

Token	Score	Description
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". Note: The NOCTM token might act as an alert to potential data entry problems. For example, an input data set address of Green St. in Raleigh has the wrong ZIP code of 27601. It is compared to the lookup data set address of Green St. in Knightdale with the correct ZIP code of 27601. The NOCTM token in the _NOTES_ variable indicates that though a ZIP code match was found, the cities did not match.
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.
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".

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

Token	Score	Description
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.

Appendix C – Zip Code Decision Hierarchy

Rather than only outputting a ZIP code, the macros assign one of these final categories to every record:

- Kept from respondent (highest protection)
    - RESPONDENT_ZIP: valid respondent ZIP; never overwritten

- From SAS geocoding / SAS-side lookups
    - SAS_MATCHED_ZIP: ZIP from PROC GEOCODE street match
    - ZIP_MATCHED_FILL: ZIP from ZIP-level geocode match
    - SINGLE_ZIP_CITY: inferred ZIP where the city has exactly one ZIP
    - MULTI_ZIP_CITY: no street and city contains multiple ZIPs (so it cannot infer a unique ZIP)

- From Python boundary lookup
    - PYTHON_GPD_FILL: ZIP assigned by GeoPandas boundary lookup (ZIP_FILL_GPD)

- Explicit exception / validation outcomes
    - INVALID_CITY_STATE: missing/non-geographic city/state
    - RESP_ZIP_BAD_STATE: respondent ZIP fails state validation (but respondent ZIP still isn't overwritten—this is documenting the issue)
    - SAS_MATCHED_ZIP_BAD_STATE: geocoder ZIP fails state validation
    - STREET_MATCH_NO_ZIP: street matched but no ZIP returned (often boundary-edge ambiguity)
    - UNCHANGED_MISSING: no usable ZIP from any source
    - FOREIGN_NO_ZIP: foreign/military address; excluded from ZIP fills

ZIP_FLAG	DESCRIPTION	HAS_ZIP	NO_ZIP	TOTAL
RESPONDENT_ZIP	Respondent-provided ZIP	18,071	0	18,071
SAS_MATCHED_ZIP	Geocoder street-level fill	8,276	0	8,276
SINGLE_ZIP_CITY	Single-ZIP city assigned	3,414	0	3,414
PYTHON_GPD_FILL	Python boundary fill	177	0	177
ZIP_INVALID[2]	Respondent entered non-ZIP value	0	15	15
STREET_MATCH_NO_ZIP	Street match, no ZIP returned	0	22	22
RESP_ZIP_BAD_STATE[3]	Respondent ZIP fails state validation	0	33	33
INVALID_CITY_STATE	Missing city or state	0	884	884
UNCHANGED_MISSING	No fill source available	0	1,048	1,048
FOREIGN_NO_ZIP	Foreign/military address	0	1,646	1,646
MULTI_ZIP_CITY	Multi-ZIP, no street	0	5,251	5,251

[2] ZIP_INVALID Respondent-provided value contained digits but did not match a valid US ZIP code pattern (common example: a year entered as ZIP) City has multiple ZIP codes so no fallback available.

[3] ZIP codes that failed state validation. The respondent-provided ZIP code did not correspond to the reported state of residence. Because the reported city has multiple ZIP codes, no unambiguous alternative could be assigned.

Appendix D – Transpose Macro Call

```
/* Step 1: Load the macro -- update path to where you saved the file */
%include "C:\your\code\folder\lhms_transpose_macro.sas";

/* Step 2: Run the conversion */
/* Output is saved temporarily to WORK.lhms15_wide by default      */

%lhms_run(
   year  = 15,                          /* 2-digit wave year            */
   indir = C:\your\data\folder\,        /* folder containing the data */
   ver   = 2                            /* data file version number   */
);
```

To save the output permanently instead of to WORK:

1. Define a library pointing to your output folder
2. Add outlib= to the %***lhms_run*** call

```
libname myout "C:\your\output\folder\";

%lhms_run(
   year   = 15,
   indir  = C:\your\data\folder\,
   ver    = 2,
   outlib = myout                       /* saves to myout.lhms15_wide */
);
```