HRS Documentation Report

Linking 1940 U.S. Census Data to the Health and Retirement Survey: Technical Documentation

John Robert Warren
Minnesota Population Center
University of Minnesota

Fabian Pfeffer
Department of Sociology
Institute for Social Research
University of Michigan

Jonas Helgertz
Minnesota Population Center
University of Minnesota
Centre for Economic Demography and Department of Economic History
Lund University

Dafeng Xu
Evans School of Public Policy & Governance
University of Washington

Version 2, October 2022

This project received support from the U.S. National Institutes of Health (NIH) via grant R01AG050300. It also benefitted from administrative and technical support from the Minnesota Population Center, which receives core infrastructure support from NIH grant 2P2CHD041023, and the Michigan Center on the Demography of Aging, which receives core infrastructure support from NIH grant P30 AG012846. The Health and Retirement Study is sponsored by the National Institute on Aging (NIA U01AG009740).

Linking 1940 U.S. Census Data to the Health and Retirement Survey: Technical Documentation

1. INTRODUCTION

In this document we describe a project to link records from the 1940 U.S. Census to records for respondents to the Health and Retirement Survey (HRS).

The project is part of a larger effort to conduct parallel linkages to the 1940 Census for respondents to the HRS, the Panel Study of Income Dynamics (PSID), the Wisconsin Longitudinal Study (WLS), the National Social Life, Health, and Aging Project (NSHAP), and the National Health and Aging Trends Study (NHATS). In each cohort study, many sample members were alive at the time of the 1940 federal census and were thus enumerated (along with their families and household members). These five ongoing longitudinal studies are central components of America's data infrastructure for interdisciplinary research on aging and the life course; physical and mental health, disability, and well-being; later-life work, economic well-being, and retirement; end-of-life issues, and many other topics. Adding information about sample members from the 1940 Census will expand the utility of all five projects and will enable important research on the effects of early life social, economic, environmental, contextual, and other factors on subsequent life outcomes.

Broadly, the project described in this document involved (1) preparing and formatting data files containing respondents' identifying information; (2) deploying machine learning algorithms to mechanically link project records to the 1940 U.S. Census; (3) hand linking records that could not be machine linked and hand-verifying a portion of those that could; and (4) documenting the new measures and making them available as part of the HRS's restricted access dissemination systems in a manner consistent with HRS respondents' privacy rights. In this document we describe the linking procedures, explain the structure of the resulting linked files and how they can be accessed, and provide information about linkage rates and the reliability and validity of the links.

This project was led by Dr. John Robert Warren, Dr. Jonas Helgertz, Dr. Dafeng Xu and others at the Minnesota Population Center (MPC) in collaboration with the Fabian Pfeffer and the HRS leadership team at the University of Michigan. It received financial support from the U.S. National Institutes of Health (NIH) via a grant (R01AG050300) to Dr. Warren. It also benefitted from administrative and technical support from MPC, which receives core infrastructure support from NIH (2P2CHD041023), and the Michigan Center on the Demography of Aging (NIH P30 AG012846). The HRS is sponsored by the National Institute on Aging (NIA U01AG009740).

1A. UPDATED ENUMERATION DISTRICT CODES IN HRS-1940 GEOGRAPHIC FILE

The ENUMDIST (enumeration district) codes available in the HRS-1940 Geographic file have been updated to reflect corrections published by IPUMS on April 19, 2021. The updated enumeration district values are now available in the ENUMV2 variable.

Prior to a April 19, 2021, correction, codes in the ENUMDIST variable were incomplete for two reasons. First, the ENUMDIST variable was disseminated as a four-digit number. In most counties, having a four-digit ENUMDIST variable sufficed, but for eight of the most populated U.S. counties (Table 1), the ENUMDIST variable should have been five digits. By truncating the extra digit, we lost many unique ENUMDISTs in these counties. The correction recovered the extra digit so all enumeration districts in these eight counties were identified.

Table 1. Eight counties with five-digit ENUMDIST values in the 1940 complete-count data.

STATEFIP	COUNTYICP	County Name
6	370	Los Angeles county, CA
17	310	Cook county, IL
26	1630	Wayne county, MI
36	50	Bronx county, NY
36	470	Kings county, NY
36	610	New York county, NY
36	810	Queens county, NY
42	1010	Philadelphia county, PA

Second, the information provided in the SUPDIST variable should have been included in the ENUMDIST codes. The correction concatenated the codes in SUPDIST and ENUMDIST into the ENUMV2 variable.

Using the new ENUMV2 variable

To use the new ENUMV2 variable, users must read in a state variable, a county variable, and the ENUMDIST variable. These three variables uniquely identify an enumeration district in the 1940 complete count data.

Prior ENUMDIST variable

For users who have been using the HRS-1940 Geographic File prior to this update, we are retaining the older, incomplete version of ENUMDIST in the ENUMDIST variable. If you plan to use the original ENUMDIST variable, users need to read in a state variable, a county variable, the SUPDIST variable, and ENUMDIST to uniquely identify 1940 enumeration district variables.

2. DATA STRUCTURE AND ACCESS

Who did we try to link?

In principle, we attempted to locate in the 1940 Census every HRS sample member who was alive by the time of the decennial enumeration on April 1, 1940. This includes the bulk of the original HRS cohort (born 1931-1941); all of the "Asset and Health Dynamics among the Oldest Old" HRS cohort (born before 1924); and all of the "Children of the Depression" HRS cohort (born 1924-1930). In all, we attempted to link 20,065 HRS respondents.

What did we produce?

The product of our lining work was a two variable file that crosswalks (1) HISTID, the permanent person-level identifier in the IPUMS.org version of the complete-count 1940 Census (Ruggles et al. 2020), and (2) the HRS identifier unique identifier.

In principle, every variable in the HRS pertaining to people we attempted to link is available to be linked to the 1940 Census data. Likewise, every variable in the publicly available complete-count 1940 Census file produced by IPUMS.org at the University of Minnesota (https://usa.ipums.org/usa/full_count.shtml) is available to be linked to HRS survey data. This includes individual level census data (for HRS sample members but also anyone they were living with in 1940), family and household level data, and geographic/ administrative information (e.g., state, county, enumeration district, street address). However, as described below, and to protect HRS sample members' identities and privacy, some information is suppressed in files available to researchers.

Structure of the Linked 1940-HRS Files

The linked 1940-HRS data are contained in three different data files: a household-level file, a person-level file, and an other person file which contains information on non-HRS household members.

1. HRS1940_Household

This file contains the household level information from the 1940 Census for those 1940 Census households who have been linked to HRS respondents. It contains 9,654 records. Variables include information on household size and structure and socio-economic circumstances (e.g., ownership of dwelling, house value). We have also added the variable HHID, containing household ID for HRS respondents. The HHID number links the HHFile with the HRS Person file as well as the Non-HRS Person File. HHID replaces the Census variables SERIAL and SERIALP to protect the anonymity of our respondents. There is a one-to-many match between the HHID and each of the Person Level files as multiple persons may reside in the same Census household.

2. HRS1940_Person

This file contains the individual level information from the 1940 census for those HRS respondents who have been linked to a 1940 Census household. It contains 9,654 records. Variables include information on relationship to the household head, age, sex, race and ethnicity, marital status, birth place, citizenship status, school enrollment and educational attainment, place of birth, place of residence in 1935, labor force status, industry, occupation, working hours, and income.

This file can be merged to the HRS 1940 Household file using the variable HHID.

3. HRS1940_Otherp

This file contains all persons who live in a 1940 Census household that has been linked to HRS respondent but who have not been identified as HRS individuals themselves. It contains 41,595 records and the same variables as included on the HRS1940 Person file.

Standard and Geographic Versions

For each type of file, there is a Standard version and a Geographic version; the version to which a researcher would have access will depend on their analytic needs. Geographic identifiers are for the most part, not available with the Standard version of the files. The Geographic version can be requested within the restricted use data contract. The 1940 Census enumeration district variable (ENUMV2) can be made available with the Geographic version of the files by special request. Due to the level of precision of enumeration districts, researchers must provide an explicit and detailed justification for exactly how and why the research will benefit from having access to this variable. Your Research Plan must include a description of the data that you will merge to the enumeration district. Details on the contents of each file are available in the Appendix. For information regarding the updated version of ENUMDIST see HRS data product page description.

How to access survey data linked to the 1940 Census

To protect respondents' identities and privacy, researchers need to access linked census data via the HRS restricted data access and licensing protocols. This means, in general, that linked files will be available remotely via a secure server at the University of Michigan's Center on the Demography of Aging (MiCDA) and only after appropriate data use agreements and security protocols have been established. The Geographic versions of the files are only available in the MiCDA enclave environment. To gain access to the Geographic versions, users will need to explain why they require access to the geographic variables and what external data sources, if any, they intend to link to the HRS data. Users are encouraged to use the Standard version in the MiCDA environment but can be made available through a traditional restricted data license if there is demonstrated need. To apply to access linked 1940-HRS data, please read the instructions here: https://hrs.isr.umich.edu/data-products/restricted-data.

3. LINKING PROCEDURES

The objective of the project—to identify HRS respondents in the 1940 full count census—was achieved using a combination of a machine learning algorithm and hand linking protocols. In this section, we present an overview of our linking procedures.

Data Cleaning

Machine linkage algorithms compare text strings and quantify their similarity; in our case, strings from the 1940 Census were compared to those from the HRS. However, formatting issues—whether strings are in upper case or lower case, whether spaces have been removed, whether punctuation is included, how abbreviations are treated, etc.—influence the results of machine record linkage. The strings "North Dakota" "N. Dakota" and "northdakota" may cognitively seem the same to a human being, but they would be treated as very different strings by a computer. So, we began by standardizing place names and given names (e.g., "Willie" and "Wm." were transformed into "William").

Machine Linking Algorithm

Our machine linkage algorithms considered PSID sample members' first names, last names, women's last names in 1940, states of birth, and years of birth (inferred from their ages) in attempting to locate corresponding individuals in the 1940 Census. As described below, other information about HRS sample members—which is available for only portions of the sample—was used in the process of hand linking and hand verification.

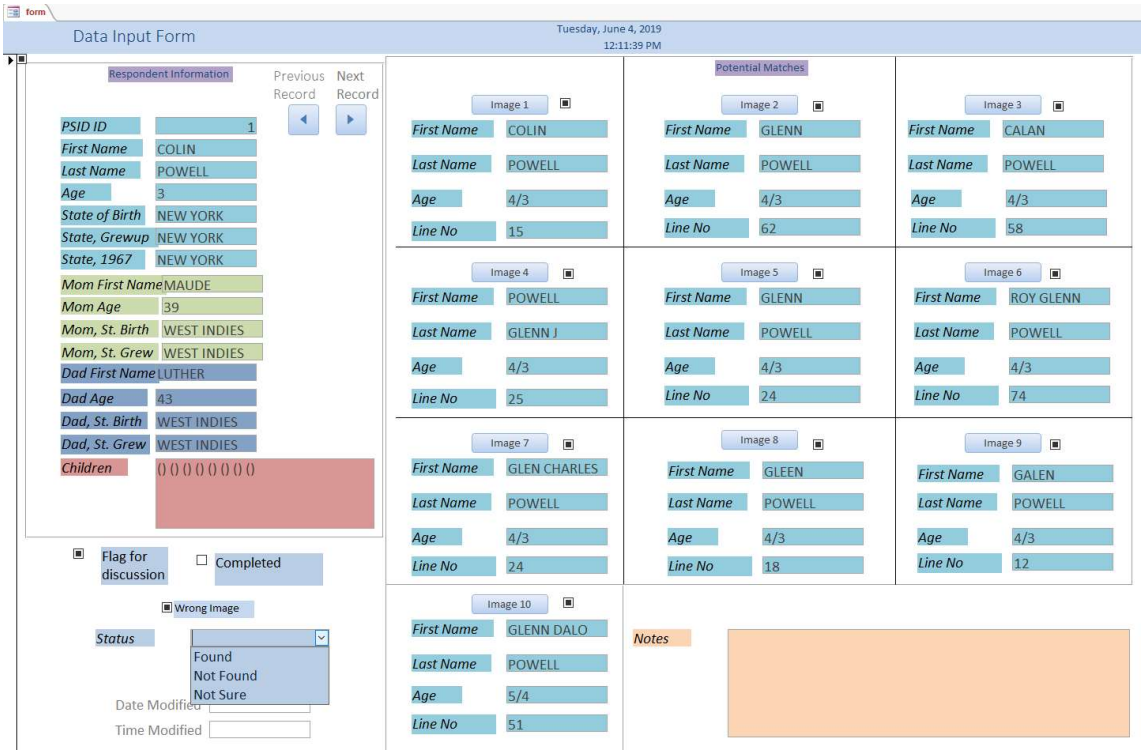
To match HRS sample members' records to the 1940 Census we first defined the universe of potential matches. To make the task computationally tractable, we restricted the population of potential matches to HRS sample members in the 1940 Census to records that displayed identical or similar characteristics on features that should be consistent over time---in this case, year of birth. For example, when attempting to find the 1940 Census record for Michael Corcoran—born in 1917 according to HRS records—we limited the population of potential matches in the 1940 Census to males born between 1914 and 1920. Because age was reported in the 1940 Censuses rather than year of birth, and because of reporting inaccuracies, we allowed for deviations of ±3 years in birth year across data sources.

To identify the correct record from among all possible records in the 1940 Census—e.g., the correct "Michael Corcoran" in 1940 from among all the possible Michael Corcorans—we employed the probabilistic linking algorithm built into the Stata utility reclink2 (Wasi and Flaaen 2015); briefly, reclink2 build on Blasnik's (2010) approach, which assigns scores to pairs of 1940-HRS records in an effort to determine which single 1940 Census record is most likely to be the hRS sample member. This probabilistic procedure typically yielded a small number of possible matches for HRS sample members in the 1940 Census. To adjudicate between multiple possible matches, we employed the hand linking and verification procedures described below. In general, we declared a linkage when there was one and only one high quality match.

Hand Linking and Verification Procedures

To verify the work of machine linking algorithm and to attempt to confidently determine high quality matches when the machine algorithm failed to do so, we next performed a set of hand linkages.

For most of the 20,065 cases, a trained hand linker was presented with a screen like the one below. On the left side is information about the focal HRS sample member (who in this fictional example is General Colin Powell). The hand linker is provided information (when available) about the HRS person's first name, last name, women's last name in 1940, age (in 1940), and state of birth. Beyond this information—which was used in the machine linking—the hand linker is also provided with information not used in machine linking, including (when available) siblings' and/or children's names, parents' years of birth, and race/ethnicity.



On the right side of the figure are up to 10 possible matches in the 1940 Census. These were the (up to 10) best matches generated by the machine linking algorithm above; hand linkers were not told which (if any) of the possible matches were declared by the machine to be a valid match. Note that hand linkers were provided with each possible match's name, age (in 1940, and a line number. The line number corresponds to the line on the 1940 Census enumeration form on which the person appears; the hand linker was then able to click on the "Image" button to view the actual 1940 Census enumeration form. Hand linkers proceeded by comparing the information about possible matches provided and on the enumeration form with information about the focal HRS sample member. Being able to see the actual enumeration forms allowed hand linkers to consider information like place of

birth, place of residence in 1940, parents' ages, and siblings' and children's names. If the hand linker determined that one—and only one—of the 10 presented options was a valid and correct match, then checked "found" and "completed" in the lower left of the screen. If they could not decide between two possible matches with the information provided, they declared "not found." In cases in which hand linkers were unable to reach a decision, they marked "not sure" and the case was reviewed by one of the authors.

To assess the reliability and validity of the overall linking efforts, we further implemented two procedures in the hand linking phase. First, we assigned hand linkers to attempt to hand link a random subset of the cases that the machine linkage algorithm declared had one and only one high-quality match. This allowed us to assess validity—how often the hand linker selected (by hand) the same match in the 1940 Census as the machine selected (via computer algorithm). Second, we assigned a randomly selected subset of hand linking cases to multiple hand linkers. This allowed us to assess the reliability of the hand linkers' decisions.

All hand linkers met regularly with the authors to ask questions; to review problematic or instructive cases; review reliability and validity statistics; and discuss progress. When necessary, hand linkers whose work was insufficiently valid or reliable were retrained or replaced.

4. LINKING RESULTS

Linkage Rates

Of the 20,065 HRS sample members we attempted to link, we were able to confidently declare high-quality matches for 9,654 (or 48.1%) of them to the 1940 Census.

Reliability and Validity

As noted above, we assessed the validity of the matches by assigning to the hand linkers a random subset of cases for which the machine algorithm declared there to be one and only one high-quality match. For these cases, the machine algorithm and the hand linker arrived at the same decision about a match 63.6% of the time. However, in most cases disagreements between the machine algorithm and the hand linker were due to the hand linker's reluctance to declare any match from among the options available to them. When hand linkers did choose a match from among the options available to them, they chose the same match as the computer algorithm 95.1% of the time.

To assess the reliability of hand linkers' efforts, we assigned a random subset of hand lining cases to two hand linkers who then independently worked the cases. For these cases, the two independent hand linkers reached the same conclusion 83.1% of the time. However, most of the disagreement in these instances was due to one linker declaring a match and the other declaring no match. When both linkers declared matches, they chose the same match 97.5% of the time.

5. IF YOU NEED TO KNOW MORE

This document is intended to serve as a brief overview to the HRS 1940 Census data products. 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.

A. 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.

B. 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 (http://hrsonline.isr.umich.edu)
E-mail: hrsquestions@umich.edu

C. Citing this Document

Please include the following citation in any research reports, papers, or publications based on these data:

In text: " This project received support from the U.S. National Institutes of Health (NIH) via grant R01AG050300 as well as infrastructure support from NIH grant 2P2CHD041023 and P30 AG012846. The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (NIA U01AG009740) and is conducted by the University of Michigan."

In references: ". Warren JR., Pfeffer F, Helgertz J, Xu D. Linking 1940 U.S. Census Data to the Health and Retirement Survey: Technical Documentation – Release 1. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan; 2020."

6. REFERENCES

Blasnik, Michael. 2010. "reclink: Stata module to probabilistically match records." Statistical Software Components S456876. Department of Economics, Boston College.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2020. " IPUMS USA: Version 10.0." Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V10.0.

Wasi, Nada, and Aaron Flaaen. 2015. "Record linkage using Stata: Preprocessing, linking, and reviewing utilities." *The STATA Journal* 15:672–97.

APPENDIX – File Contents

HRS1940_Household - Standard Version

Variable	Type	Len	Format	Label
HHIDPN	Char	9		Household Identifier and Person Number
HHID	Char	6	$6.	HRS household identifier
YEAR	Num	8		Census year
NUMPREC	Num	8		Number of person records following
SUBSAMP	Num	8		Subsample number
DWSIZE	Num	8		Dwelling size
REGION	Num	8		Census region and division
SIZEPL	Num	8	Z2.	Size of place
URBAN	Num	8		Urban/rural status
GQ	Num	8		Group quarters status
GQTYPE	Num	8		Group quarters type
GQFUNDS	Num	8		Group quarters funding
FARM	Num	8		Farm status
OWNERSHP	Num	8		Ownership of dwelling (tenure)
NFAMS	Num	8		Number of families in household
NCOUPLES	Num	8		Number of couples in household
NMOTHERS	Num	8		Number of mothers in household
NFATHERS	Num	8		Number of fathers in household
QFARM	Num	8		Flag for Farm
QOWNERSH	Num	8		Flag for Ownershp
HHTYPE	Num	8		Household Type
NSUBFAM	Num	8		Number of subfamilies in household
HEADLOC	Num	8		Location of household head
VALUEH	Num	8		House value
MULTGEN	Num	8		Multigenerational household
CPI99	Num	8		CPI-U adjustment factor to 1999 dollars
HHWT	Num	8		Household weight
DWSEQ	Num	8		Household sequence within dwelling
RENT	Num	8		Monthly contract rent
RESPOND	Num	8		Respondent's relationship to household head
NUMPERHH	Num	8		Number of persons in household
SLPERNUM	Num	8		Sample-line person number
QGQFUNDS	Num	8		Flag for Gqfunds
SPLIT40	Num	8		Large group quarters that was split up, 1940 100%
EDMISS	Num	8		Identifies households in missing data enumeration districts
RECTYPEP	Char	1		Record type

HRS1940_Person - Standard Version

Variable	Type	Len	Format	Label
HHIDPN	Char	9		Household Identifier and Person Number
HHID	Char	6	$6.	HRS household identifier
RECTYPE	Char	1		Record type
YEARP	Num	8		Census year
PERNUM	Num	8		Person number in sample unit
SLWTREG	Num	8		Sample-line weight (integral)
MOMLOC	Num	8		Mother's location in the household
STEPMOM	Num	8		Probable step/adopted mother
MOMRULE_HIST	Num	8		Rule for linking mother
POPLOC	Num	8		Father's location in the household
STEPPOP	Num	8		Probable step/adopted father
POPRULE_HIST	Num	8		Rule for linking father
SPLOC	Num	8		Spouse's location in household
SPRULE_HIST	Num	8		Rule for linking spouse
FAMSIZE	Num	8		Number of own family members in household
NCHILD	Num	8		Number of own children in the household
NCHLT5	Num	8		Number of own children under age 5 in household
FAMUNIT	Num	8		Family unit membership
ELDCH	Num	8		Age of eldest own child in household
YNGCH	Num	8		Age of youngest own child in household
NSIBS	Num	8		Number of own siblings in household
RELATE	Num	8		Relationship to household head
AGE	Num	8		Age
SEX	Num	8		Sex
RACE	Num	8		Race
MARST	Num	8		Marital status
MARRNO	Num	8		Times married
CHBORN	Num	8		Children ever born
SLREC	Num	8		Sample-line person identifier
BPL	Num	8		Birthplace
NATIVITY	Num	8		Foreign birthplace or parentage
CITIZEN	Num	8		Citizenship status
HISPAN	Num	8		Hispanic origin
MTONGUE	Num	8		Mother tongue
SPANNAME	Num	8		Spanish surname
SCHOOL	Num	8		School attendance
HIGRADE	Num	8		Highest grade of schooling
EMPSTAT	Num	8		Employment status
LABFORCE	Num	8		Labor force status
OCC1950	Num	8		Occupation, 1950 basis
OCCSCORE	Num	8		Occupational income score
SEI	Num	8		Duncan Socioeconomic Index
IND1950	Num	8		Industry, 1950 basis
CLASSWKR	Num	8		Class of worker
WKSWORK1	Num	8		Weeks worked last year
WKSWORK2	Num	8		Weeks worked last year, intervalled
HRSWORK1	Num	8		Hours worked last week
HRSWORK2	Num	8		Hours worked last week, intervalled
INCWAGE	Num	8		Wage and salary income
INCNONWG	Num	8		Had non-wage/salary income over $50
QAGE	Num	8		Flag for Age
QBPL	Num	8		Flag for Bpl, Nativity
QCHBORN	Num	8		Flag for Chborn
QCITIZEN	Num	8		Flag for Citizen
QCLASSWK	Num	8		Flag for Classwkr
QFBPL	Num	8		Flag for Fbpl, Nativity
QEMPSTAT	Num	8		Flag for Empstat, Labforce

Variable	Type	Len	Format	Label
QMARST	Num	8		Flag for Marst
QMTONGUE	Num	8		Flag for Mtongue
QOCC	Num	8		Flag for Occ, Occ1950, SEI, Occscore, Occsoc, Labforce
QRACE	Num	8		Flag for Race
QRELATE	Num	8		Flag for Relate
QSURSIM	Num	8		Flag for Sursim
QSCHOOL	Num	8		Flag for School, Schltype
QSEX	Num	8		Flag for Sex
QYRIMM	Num	8		Flag for Yrimmig, Yrsusa1, Yrsusa2
AGEDIFF	Num	8		Temporary
HISPRULE	Num	8		Hispanic origin rule
PRESGL	Num	8		Occupational prestige score, Siegel
ERSCOR50	Num	8		Occupational earnings score, 1950 basis
EDSCOR50	Num	8		Occupational education score, 1950 basis
NPBOSS50	Num	8		Nam-Powers-Boyd occupational status score, 1950 basis
ISRELATE	Num	8		[relate flag]
SUBFAM	Num	8		Subfamily membership
SFTYPE	Num	8		Subfamily type
SFRELATE	Num	8		Relationship within subfamily
EDUC	Num	8		Educational attainment
VETSTAT	Num	8		Veteran status
IND	Num	8		Industry
SLWT	Num	8		Sample-line weight
PERWT	Num	8		Person weight
BIRTHYR	Num	8		Year of birth
OCC	Num	8		Occupation
MIGRATE5	Num	8		Migration status, 5 years
MIGPLAC5	Num	8		State or country of residence 5 years ago
MIGTYPE5	Num	8		Metropolitan status 5 years ago
MIGFARM5	Num	8		Farm status 5 years ago
AGEMARR	Num	8		Age at first marriage
MBPL	Num	8		Mother's birthplace
FBPL	Num	8		Father's birthplace
DURUNEMP	Num	8		Continuous weeks unemployed
UCLASSWK	Num	8		Usual class of worker
UOCC	Num	8		Usual occupation
UOCC95	Num	8		Usual occupation, 1950 classification
UIND	Num	8		Usual industry
SAMEPLAC	Num	8		Lived same incorporated place 5 years ago
SAMEMET5	Num	8		Lived same metropolitan area 5 years ago
SAMESEA5	Num	8		Lived same SEA 5 years ago
AGEMONTH	Num	8		Age in months
RESPONDT	Num	8		Respondent indicator
VETWWI	Num	8		Veteran, served during WWI era
VET1940	Num	8		Veteran status, 1940
VETPER	Num	8		Veteran period of service, 1940
VETCHILD	Num	8		Mortality status of child's veteran father
SURSIM	Num	8		Surname similarity
SSENROLL	Num	8		Social Security enrollment
QMBPL	Num	8		Flag for Mbpl, Nativity
OCC1940	Char	3		Occupation, Unrecoded 1940
IND1940	Char	3		Industry, Unrecoded 1940
ANYALLOCATION	Num	8		Record had any value allocated

HRS1940_Otherp - Standard Version

Variable	Type	Len	Format	Label
HHIDPN	Char	9		Household Identifier and Person Number
HHID	Char	6	$6.	HRS household identifier
RECTYPE	Char	1		Record type
YEARP	Num	8		Census year
PERNUM	Num	8		Person number in sample unit
SLWTREG	Num	8		Sample-line weight (integral)
MOMLOC	Num	8		Mother's location in the household
STEPMOM	Num	8		Probable step/adopted mother
MOMRULE_HIST	Num	8		Rule for linking mother
POPLOC	Num	8		Father's location in the household
STEPPOP	Num	8		Probable step/adopted father
POPRULE_HIST	Num	8		Rule for linking father
SPLOC	Num	8		Spouse's location in household
SPRULE_HIST	Num	8		Rule for linking spouse
FAMSIZE	Num	8		Number of own family members in household
NCHILD	Num	8		Number of own children in the household
NCHLT5	Num	8		Number of own children under age 5 in household
FAMUNIT	Num	8		Family unit membership
ELDCH	Num	8		Age of eldest own child in household
YNGCH	Num	8		Age of youngest own child in household
NSIBS	Num	8		Number of own siblings in household
RELATE	Num	8		Relationship to household head
AGE	Num	8		Age
SEX	Num	8		Sex
RACE	Num	8		Race
MARST	Num	8		Marital status
MARRNO	Num	8		Times married
CHBORN	Num	8		Children ever born
SLREC	Num	8		Sample-line person identifier
BPL	Num	8		Birthplace
NATIVITY	Num	8		Foreign birthplace or parentage
CITIZEN	Num	8		Citizenship status
HISPAN	Num	8		Hispanic origin
MTONGUE	Num	8		Mother tongue
SPANNAME	Num	8		Spanish surname
SCHOOL	Num	8		School attendance
HIGRADE	Num	8		Highest grade of schooling
EMPSTAT	Num	8		Employment status
LABFORCE	Num	8		Labor force status
OCC1950	Num	8		Occupation, 1950 basis
OCCSCORE	Num	8		Occupational income score
SEI	Num	8		Duncan Socioeconomic Index
IND1950	Num	8		Industry, 1950 basis
CLASSWKR	Num	8		Class of worker
WKSWORK1	Num	8		Weeks worked last year
WKSWORK2	Num	8		Weeks worked last year, intervalled
HRSWORK1	Num	8		Hours worked last week
HRSWORK2	Num	8		Hours worked last week, intervalled
INCWAGE	Num	8		Wage and salary income
INCNONWG	Num	8		Had non-wage/salary income over $50
QAGE	Num	8		Flag for Age
QBPL	Num	8		Flag for Bpl, Nativity
QCHBORN	Num	8		Flag for Chborn
QCITIZEN	Num	8		Flag for Citizen
QCLASSWK	Num	8		Flag for Classwkr
QFBPL	Num	8		Flag for Fbpl, Nativity
QEMPSTAT	Num	8		Flag for Empstat, Labforce

Variable	Type	Len	Format	Label
QMARST	Num	8		Flag for Marst
QMTONGUE	Num	8		Flag for Mtongue
QOCC	Num	8		Flag for Occ, Occ1950, SEI, Occscore, Occsoc, Labforce
QRACE	Num	8		Flag for Race
QRELATE	Num	8		Flag for Relate
QSURSIM	Num	8		Flag for Sursim
QSCHOOL	Num	8		Flag for School, Schltype
QSEX	Num	8		Flag for Sex
QYRIMM	Num	8		Flag for Yrimmig, Yrsusa1, Yrsusa2
AGEDIFF	Num	8		Temporary
HISPRULE	Num	8		Hispanic origin rule
PRESGL	Num	8		Occupational prestige score, Siegel
ERSCOR50	Num	8		Occupational earnings score, 1950 basis
EDSCOR50	Num	8		Occupational education score, 1950 basis
NPBOSS50	Num	8		Nam-Powers-Boyd occupational status score, 1950 basis
ISRELATE	Num	8		[relate flag]
SUBFAM	Num	8		Subfamily membership
SFTYPE	Num	8		Subfamily type
SFRELATE	Num	8		Relationship within subfamily
EDUC	Num	8		Educational attainment
VETSTAT	Num	8		Veteran status
IND	Num	8		Industry
SLWT	Num	8		Sample-line weight
PERWT	Num	8		Person weight
BIRTHYR	Num	8		Year of birth
OCC	Num	8		Occupation
MIGRATE5	Num	8		Migration status, 5 years
MIGTYPE5	Num	8		Metropolitan status 5 years ago
MIGFARM5	Num	8		Farm status 5 years ago
AGEMARR	Num	8		Age at first marriage
MBPL	Num	8		Mother's birthplace
FBPL	Num	8		Father's birthplace
DURUNEMP	Num	8		Continuous weeks unemployed
UCLASSWK	Num	8		Usual class of worker
UOCC	Num	8		Usual occupation
UOCC95	Num	8		Usual occupation, 1950 classification
UIND	Num	8		Usual industry
SAMEPLAC	Num	8		Lived same incorporated place 5 years ago
SAMEMET5	Num	8		Lived same metropolitan area 5 years ago
SAMESEA5	Num	8		Lived same SEA 5 years ago
AGEMONTH	Num	8		Age in months
RESPONDT	Num	8		Respondent indicator
VETWWI	Num	8		Veteran, served during WWI era
VET1940	Num	8		Veteran status, 1940
VETPER	Num	8		Veteran period of service, 1940
VETCHILD	Num	8		Mortality status of child's veteran father
SURSIM	Num	8		Surname similarity
SSENROLL	Num	8		Social Security enrollment
QMBPL	Num	8		Flag for Mbpl, Nativity
OCC1940	Char	3		Occupation, Unrecoded 1940
IND1940	Char	3		Industry, Unrecoded 1940
ANYALLOCATION	Num	8		Record had any value allocated

*HRS1940_Household - Geographic Version**

Variable	Type	Len	Format	Label
HHIDPN	Char	9		Household Identifier and Person Number
HHID	Char	6	$6.	HRS household identifier
YEAR	Num	8		Census year
NUMPREC	Num	8		Number of person records following
SUBSAMP	Num	8		Subsample number
DWSIZE	Num	8		Dwelling size
REGION	Num	8		Census region and division
STATEICP	Num	8		State (ICPSR code)
STATEFIP	Num	8		State (FIPS code)
SEA	Num	8		State Economic Area
METRO	Num	8		Metropolitan status
METAREA	Num	8		Metropolitan area
METDIST	Num	8		Metropolitan district
CITY	Num	8		City
CITYPOP	Num	8		City population
SIZEPL	Num	8	Z2.	Size of place
URBAN	Num	8		Urban/rural status
GQ	Num	8		Group quarters status
GQTYPE	Num	8		Group quarters type
GQFUNDS	Num	8		Group quarters funding
FARM	Num	8		Farm status
OWNERSHP	Num	8		Ownership of dwelling (tenure)
NFAMS	Num	8		Number of families in household
NCOUPLES	Num	8		Number of couples in household
NMOTHERS	Num	8		Number of mothers in household
NFATHERS	Num	8		Number of fathers in household
QFARM	Num	8		Flag for Farm
QOWNERSH	Num	8		Flag for Ownershp
URBPOP	Num	8		Population of urban places
HHTYPE	Num	8		Household Type
CNTRY	Num	8		Country
NSUBFAM	Num	8		Number of subfamilies in household
HEADLOC	Num	8		Location of household head
VALUEH	Num	8		House value
MULTGEN	Num	8		Multigenerational household
CPI99	Num	8		CPI-U adjustment factor to 1999 dollars
COUNTYICP	Num	8		County (ICPSR code)
HHWT	Num	8		Household weight
DWSEQ	Num	8		Household sequence within dwelling
RENT	Num	8		Monthly contract rent
RESPOND	Num	8		Respondent's relationship to household head
NUMPERHH	Num	8		Number of persons in household
SLPERNUM	Num	8		Sample-line person number
SUPDIST	Num	8		Supervisor's district number
QGQFUNDS	Num	8		Flag for Gqfunds
SPLIT40	Num	8		Large group quarters that was split up, 1940 100%
NUMPREC40	Num	8		Number of person records in household, before large group quarters were split up, 1940 100%
EDMISS	Num	8		Identifies households in missing data enumeration districts
SPLIT	Num	8		Large group quarters that was split up (100% datasets)
SPLITNUM	Num	8		Number of person records in household, before large group quarters were split up (100% datasets)
RECTYPEP	Char	1		Record type

* ENUMV2 and the original ENUMDIST (Enumeration district) can be added to this file by request

HRS1940_Person - Geographic Version

Variable	Type	Len	Format	Label
HHIDPN	Char	9		Household Identifier and Person Number
HHID	Char	6	$6.	HRS household identifier
RECTYPE	Char	1		Record type
YEARP	Num	8		Census year
PERNUM	Num	8		Person number in sample unit
SLWTREG	Num	8		Sample-line weight (integral)
MOMLOC	Num	8		Mother's location in the household
STEPMOM	Num	8		Probable step/adopted mother
MOMRULE_HIST	Num	8		Rule for linking mother
POPLOC	Num	8		Father's location in the household
STEPPOP	Num	8		Probable step/adopted father
POPRULE_HIST	Num	8		Rule for linking father
SPLOC	Num	8		Spouse's location in household
SPRULE_HIST	Num	8		Rule for linking spouse
FAMSIZE	Num	8		Number of own family members in household
NCHILD	Num	8		Number of own children in the household
NCHLT5	Num	8		Number of own children under age 5 in household
FAMUNIT	Num	8		Family unit membership
ELDCH	Num	8		Age of eldest own child in household
YNGCH	Num	8		Age of youngest own child in household
NSIBS	Num	8		Number of own siblings in household
RELATE	Num	8		Relationship to household head
AGE	Num	8		Age
SEX	Num	8		Sex
RACE	Num	8		Race
MARST	Num	8		Marital status
MARRNO	Num	8		Times married
CHBORN	Num	8		Children ever born
SLREC	Num	8		Sample-line person identifier
BPL	Num	8		Birthplace
NATIVITY	Num	8		Foreign birthplace or parentage
CITIZEN	Num	8		Citizenship status
HISPAN	Num	8		Hispanic origin
MTONGUE	Num	8		Mother tongue
SPANNAME	Num	8		Spanish surname
SCHOOL	Num	8		School attendance
HIGRADE	Num	8		Highest grade of schooling
EMPSTAT	Num	8		Employment status
LABFORCE	Num	8		Labor force status
OCC1950	Num	8		Occupation, 1950 basis
OCCSCORE	Num	8		Occupational income score
SEI	Num	8		Duncan Socioeconomic Index
IND1950	Num	8		Industry, 1950 basis
CLASSWKR	Num	8		Class of worker
WKSWORK1	Num	8		Weeks worked last year
WKSWORK2	Num	8		Weeks worked last year, intervalled
HRSWORK1	Num	8		Hours worked last week
HRSWORK2	Num	8		Hours worked last week, intervalled
INCWAGE	Num	8		Wage and salary income
INCNONWG	Num	8		Had non-wage/salary income over $50
MIGCITY5	Num	8		City of residence 5 years ago
QAGE	Num	8		Flag for Age
QBPL	Num	8		Flag for Bpl, Nativity
QCHBORN	Num	8		Flag for Chborn
QCITIZEN	Num	8		Flag for Citizen
QCLASSWK	Num	8		Flag for Classwkr
QFBPL	Num	8		Flag for Fbpl, Nativity

Variable	Type	Len	Format	Label
QEMPSTAT	Num	8		Flag for Empstat, Labforce
QMARST	Num	8		Flag for Marst
QMTONGUE	Num	8		Flag for Mtongue
QOCC	Num	8		Flag for Occ, Occ1950, SEI, Occscore, Occsoc, Labforce
QRACE	Num	8		Flag for Race
QRELATE	Num	8		Flag for Relate
QSURSIM	Num	8		Flag for Sursim
QSCHOOL	Num	8		Flag for School, Schltype
QSEX	Num	8		Flag for Sex
QYRIMM	Num	8		Flag for Yrimmig, Yrsusa1, Yrsusa2
AGEDIFF	Num	8		Temporary
HISPRULE	Num	8		Hispanic origin rule
PRESGL	Num	8		Occupational prestige score, Siegel
ERSCOR50	Num	8		Occupational earnings score, 1950 basis
EDSCOR50	Num	8		Occupational education score, 1950 basis
NPBOSS50	Num	8		Nam-Powers-Boyd occupational status score, 1950 basis
ISRELATE	Num	8		[relate flag]
SUBFAM	Num	8		Subfamily membership
SFTYPE	Num	8		Subfamily type
SFRELATE	Num	8		Relationship within subfamily
EDUC	Num	8		Educational attainment
VETSTAT	Num	8		Veteran status
IND	Num	8		Industry
SLWT	Num	8		Sample-line weight
PERWT	Num	8		Person weight
BIRTHYR	Num	8		Year of birth
OCC	Num	8		Occupation
MIGRATE5	Num	8		Migration status, 5 years
MIGPLAC5	Num	8		State or country of residence 5 years ago
MIGMET5	Num	8		Metropolitan area of residence 5 years ago
MIGTYPE5	Num	8		Metropolitan status 5 years ago
MIGSEA5	Num	8		SEA of residence 5 years ago
MIGFARM5	Num	8		Farm status 5 years ago
AGEMARR	Num	8		Age at first marriage
MBPL	Num	8		Mother's birthplace
FBPL	Num	8		Father's birthplace
DURUNEMP	Num	8		Continuous weeks unemployed
UCLASSWK	Num	8		Usual class of worker
UOCC	Num	8		Usual occupation
UOCC95	Num	8		Usual occupation, 1950 classification
UIND	Num	8		Usual industry
SAMEPLAC	Num	8		Lived same incorporated place 5 years ago
SAMEMET5	Num	8		Lived same metropolitan area 5 years ago
SAMESEA5	Num	8		Lived same SEA 5 years ago
AGEMONTH	Num	8		Age in months
RESPONDT	Num	8		Respondent indicator
VETWWI	Num	8		Veteran, served during WWI era
VET1940	Num	8		Veteran status, 1940
VETPER	Num	8		Veteran period of service, 1940
VETCHILD	Num	8		Mortality status of child's veteran father
SURSIM	Num	8		Surname similarity
SSENROLL	Num	8		Social Security enrollment
QMBPL	Num	8		Flag for Mbpl, Nativity
OCC1940	Char	3		Occupation, Unrecoded 1940
IND1940	Char	3		Industry, Unrecoded 1940
MIGCOUNTY	Num	8		County of residence 5 years ago
ANYALLOCATION	Num	8		Record had any value allocated

HRS1940_Otherp - Geographic Version

Variable	Type	Len	Format	Label
HHIDPN	Char	9		Household Identifier and Person Number
HHID	Char	6	$6.	HRS household identifier
RECTYPE	Char	1		Record type
YEARP	Num	8		Census year
PERNUM	Num	8		Person number in sample unit
SLWTREG	Num	8		Sample-line weight (integral)
MOMLOC	Num	8		Mother's location in the household
STEPMOM	Num	8		Probable step/adopted mother
MOMRULE_HIST	Num	8		Rule for linking mother
POPLOC	Num	8		Father's location in the household
STEPPOP	Num	8		Probable step/adopted father
POPRULE_HIST	Num	8		Rule for linking father
SPLOC	Num	8		Spouse's location in household
SPRULE_HIST	Num	8		Rule for linking spouse
FAMSIZE	Num	8		Number of own family members in household
NCHILD	Num	8		Number of own children in the household
NCHLT5	Num	8		Number of own children under age 5 in household
FAMUNIT	Num	8		Family unit membership
ELDCH	Num	8		Age of eldest own child in household
YNGCH	Num	8		Age of youngest own child in household
NSIBS	Num	8		Number of own siblings in household
RELATE	Num	8		Relationship to household head
AGE	Num	8		Age
SEX	Num	8		Sex
RACE	Num	8		Race
MARST	Num	8		Marital status
MARRNO	Num	8		Times married
CHBORN	Num	8		Children ever born
SLREC	Num	8		Sample-line person identifier
BPL	Num	8		Birthplace
NATIVITY	Num	8		Foreign birthplace or parentage
CITIZEN	Num	8		Citizenship status
HISPAN	Num	8		Hispanic origin
MTONGUE	Num	8		Mother tongue
SPANNAME	Num	8		Spanish surname
SCHOOL	Num	8		School attendance
HIGRADE	Num	8		Highest grade of schooling
EMPSTAT	Num	8		Employment status
LABFORCE	Num	8		Labor force status
OCC1950	Num	8		Occupation, 1950 basis
OCCSCORE	Num	8		Occupational income score
SEI	Num	8		Duncan Socioeconomic Index
IND1950	Num	8		Industry, 1950 basis
CLASSWKR	Num	8		Class of worker
WKSWORK1	Num	8		Weeks worked last year
WKSWORK2	Num	8		Weeks worked last year, intervalled
HRSWORK1	Num	8		Hours worked last week
HRSWORK2	Num	8		Hours worked last week, intervalled
INCWAGE	Num	8		Wage and salary income
INCNONWG	Num	8		Had non-wage/salary income over $50
MIGCITY5	Num	8		City of residence 5 years ago
QAGE	Num	8		Flag for Age
QBPL	Num	8		Flag for Bpl, Nativity
QCHBORN	Num	8		Flag for Chborn
QCITIZEN	Num	8		Flag for Citizen
QCLASSWK	Num	8		Flag for Classwkr
QFBPL	Num	8		Flag for Fbpl, Nativity
QEMPSTAT	Num	8		Flag for Empstat, Labforce

Variable	Type	Len	Format	Label
QMARST	Num	8		Flag for Marst
QMTONGUE	Num	8		Flag for Mtongue
QOCC	Num	8		Flag for Occ, Occ1950, SEI, Occscore, Occsoc, Labforce
QRACE	Num	8		Flag for Race
QRELATE	Num	8		Flag for Relate
QSURSIM	Num	8		Flag for Sursim
QSCHOOL	Num	8		Flag for School, Schltype
QSEX	Num	8		Flag for Sex
QYRIMM	Num	8		Flag for Yrimmig, Yrsusa1, Yrsusa2
AGEDIFF	Num	8		Temporary
HISPRULE	Num	8		Hispanic origin rule
PRESGL	Num	8		Occupational prestige score, Siegel
ERSCOR50	Num	8		Occupational earnings score, 1950 basis
EDSCOR50	Num	8		Occupational education score, 1950 basis
NPBOSS50	Num	8		Nam-Powers-Boyd occupational status score, 1950 basis
ISRELATE	Num	8		[relate flag]
SUBFAM	Num	8		Subfamily membership
SFTYPE	Num	8		Subfamily type
SFRELATE	Num	8		Relationship within subfamily
EDUC	Num	8		Educational attainment
VETSTAT	Num	8		Veteran status
IND	Num	8		Industry
SLWT	Num	8		Sample-line weight
PERWT	Num	8		Person weight
BIRTHYR	Num	8		Year of birth
OCC	Num	8		Occupation
MIGRATE5	Num	8		Migration status, 5 years
MIGPLAC5	Num	8		State or country of residence 5 years ago
MIGMET5	Num	8		Metropolitan area of residence 5 years ago
MIGTYPE5	Num	8		Metropolitan status 5 years ago
MIGSEA5	Num	8		SEA of residence 5 years ago
MIGFARM5	Num	8		Farm status 5 years ago
AGEMARR	Num	8		Age at first marriage
MBPL	Num	8		Mother's birthplace
FBPL	Num	8		Father's birthplace
DURUNEMP	Num	8		Continuous weeks unemployed
UCLASSWK	Num	8		Usual class of worker
UOCC	Num	8		Usual occupation
UOCC95	Num	8		Usual occupation, 1950 classification
UIND	Num	8		Usual industry
SAMEPLAC	Num	8		Lived same incorporated place 5 years ago
SAMEMET5	Num	8		Lived same metropolitan area 5 years ago
SAMESEA5	Num	8		Lived same SEA 5 years ago
AGEMONTH	Num	8		Age in months
RESPONDT	Num	8		Respondent indicator
VETWWI	Num	8		Veteran, served during WWI era
VET1940	Num	8		Veteran status, 1940
VETPER	Num	8		Veteran period of service, 1940
VETCHILD	Num	8		Mortality status of child's veteran father
SURSIM	Num	8		Surname similarity
SSENROLL	Num	8		Social Security enrollment
QMBPL	Num	8		Flag for Mbpl, Nativity
OCC1940	Char	3		Occupation, Unrecoded 1940
IND1940	Char	3		Industry, Unrecoded 1940
MIGCOUNTY	Num	8		County of residence 5 years ago
ANYALLOCATION	Num	8		Record had any value allocated