

# HEALTH AND RETIREMENT STUDY

## **Restricted Data**

### **Social Security Weights**

Data Description and Usage  
(November 2016 Revision)

Version 1.1, February 2016  
*(Documentation version 3.1, June 2017)*

**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](mailto:hrsquestions@umich.edu)).

This document may not be reproduced without the written consent of the staff of the Health and Retirement Study, The Institute for Social Research, The University of Michigan.

## Table of Contents

<b>1. Overview and Acknowledgments.....</b>	<b>2</b>
<b>2. Obtaining the Data.....</b>	<b>2</b>
2a. Access to Restricted Data.....	2
2b. Restricted Data Agreement .....	2
2c. Publications Based on Restricted Data.....	2
<b>3. Data File Contents .....</b>	<b>2</b>
<b>4. Applying SS Weights .....</b>	<b>3</b>
Summary of SS weight usage .....	3
<b>5. If You Need to Know More .....</b>	<b>3</b>
5a. HRS Internet Site.....	3
5b. Contact Information .....	4
<b>APPENDIX A: Weight Calculations .....</b>	<b>5</b>
A1. Household Level Weights .....	5
Table III.H.1A. SSA Earnings Matches for 1992 Wave (Number of Households) .....	6
Table III.H.1B SSA Earnings Matches for 1998 Wave (Number of Households) .....	6
Table III.H.1C. SSA Earnings Matches for 1998 Wave (Number of Households), by entry cohort.....	6
Table III.H.1D.SSA Earnings Matches for 2004 Wave (Number of Households) .....	6
Table III.H.1E. SSA Earnings Matches for 2004 Wave (Number of Households), by entry cohort.....	7
A2. Respondent Level Weights.....	7
Table III.H.2A. SSA Earnings Matches for 1992 Wave (Number of Respondents) .....	7
Table III.H.2B. SSA Earnings Matches for 1998 Wave (Number of Respondents) .....	8
Table III.H.2C. SSA Earnings Matches for 1998 Wave, by birth cohort .....	8
<b>Appendix B: Installation (Traditional License Only).....</b>	<b>9</b>
B1. Distribution Set.....	9
B1-1. Windows Environment.....	9
B1-2. UNIX Environment .....	9
B1-3. Macintosh OS X Environment (10.4.x and above) .....	9
B1. Program Statements .....	9
B2-1. Using the Files with SAS .....	9
B2-2. Using the Files with SPSS.....	9
B2-3. Using the Files with Stata.....	9
Table B1: Contents of the Distribution Package.....	10

## 1. Overview and Acknowledgments

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 (U01AG009740), with supplemental support from the Social Security Administration. The study is conducted by the Institute for Social Research (ISR) at the University of Michigan.

## 2. Obtaining the Data

### 2a. Access to Restricted Data

Although most HRS data sets are available to the public without restriction, certain HRS data sets contain sensitive respondent information and are only available under terms of a formal agreement negotiated between the researcher and HRS. Prospective users of HRS restricted geocode data have two access options:

- [MiCDA Enclave](#) Virtual Desktop Infrastructure (recommended)
- Traditional Licensing Agreement (deprecated)

For instructions on how to proceed, visit the [HRS Restricted Data](#) page or contact the HRS Restricted Data Applications Processing Team ([hrrdaapplications@umich.edu](mailto:hrrdaapplications@umich.edu)) by email.

### 2b. Restricted Data Agreement

This restricted data set is intended for exclusive use by you and the persons specified in the *Confidentiality Agreement for Use of Restricted Data from the Health and Retirement Study*. Traditional licensing agreements that include multiple users are also bound by the *Supplemental Agreement with Research Staff for Use of Restricted Data from the Health and Retirement Study*.

### 2c. Publications Based on Restricted Data

Your restricted data agreement specifies that you will inform HRS of any papers, publications, or presentations based on this restricted data set. Please send a copy of such publications in PDF format via e-mail to [hrrquestions@umich.edu](mailto:hrrquestions@umich.edu) with "Attn: Papers and Publications" in the subject line. If you wish, you may include a bibliographical reference.

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

Health and Retirement Study  
Attn: Papers and Publications  
The Institute for Social Research, Room 3410  
P.O. Box 1248  
Ann Arbor, Michigan 48106-1248

## 3. Data File Contents

This release of the *Social Security Weights* data set is comprised of three respondent-level, cross-sectional files, each containing one set of SS weights: 1992 (n=12651), 1998 (n=22879), and 2004 (n=21820). The naming convention used is SSWGTSw\_R.\*, where A\_R designates respondent level data for 1992, F\_R designates data for 1998, and K\_R designates data for 2004. Respondents are uniquely identified by the concatenation of the household ID and the person number, HHID and PN. Refer to Appendix A of this document or to *Prospective Social Security Wealth Measures: Technical Documentation* for further information on how these weights were calculated.

## 4. Applying SS Weights

Users should keep in mind that:

- Social Security weights pertain to earnings records, not the benefits records. However, because the correlation between consent to link to SS earnings records and consent to link to SS benefits records is high in HRS, it is acceptable to use SS weights for the benefits records.
- The Social Security weights apply to files created using the earnings records for the matched sample (i.e. those whose earnings records are available) without any imputation for the unmatched sample.
- If a user needs to utilize both matched and unmatched samples in any given file, the respondent weight in the tracker file should be used.
- SS weights should not be applied to data derived from deceased spouse administrative records since no appropriate weights have been generated for this category.

Summary of SS weight usage	
Product	Use SS Weights?
Respondent Cross-Year Benefits	Yes <sup>1</sup>
Respondent Cross-Year Summary Earnings	Yes
Respondent Cross-Year Detail Earnings	Yes
Deceased Spouse Cross-Year Benefits	No
Deceased Spouse Cross-Year Summary Earnings	No
Deceased Spouse Cross-Year Detail Earnings	No
Form 831 Disability Records [Respondent; Deceased Spouse]	No <sup>2</sup>
Supplemental Security Income [Respondent; Deceased Spouse]	Yes <sup>3</sup>
Prospective Social Security Wealth Measures	Yes <sup>4</sup>

## 5. If You Need to Know More

This document is intended to serve as a brief overview of the *Social Security Weights* data set. If you have questions or concerns that are not adequately covered here or on our Web site, or if you have any comments, please contact us. We will do our best to provide answers.

### 5a. HRS Internet Site

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, point your Web browser to the HRS Web site at <http://hrsonline.isr.umich.edu/>. Visit [the HRS restricted data pages](#) to find out more about restricted data products and procedures.

---

<sup>1</sup> Since consent to link to SS earnings records and benefits records is highly correlated, using SS weights with the R Benefits file is acceptable.

<sup>2</sup> The construction of SS weights does not adjust for the likelihood of DI application

<sup>3</sup> For respondent data only, not deceased spouse.

<sup>4</sup> For analyses that are restricted to the matched sample, for which Social Security wealth is estimated from the SS linked data, the SS weight should be used. For analysis that uses imputed SS wealth data for the non-matched sample in combination with data for the matched sample, the HRS respondent weight in the Cross-wave Tracker file should be used.

## **5b. Contact Information**

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

Internet: [Help Desk](#) at our Web site

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

Postal Service:

Health and Retirement Study  
The Institute for Social Research, Room 3050  
The University of Michigan  
P.O. Box 1248  
Ann Arbor, MI 48106-1248

## APPENDIX A: Weight Calculations<sup>5</sup>

Approximately 80 percent of the HRS cohort interviewed in 1992 granted permission to HRS to obtain earnings records and benefit entitlement data from the SSA. Previous researchers have examined whether this “matched” sample is random and the evidence suggests matched and unmatched respondents vary on several observable dimensions including gender, race, education, geographic location, labor force participation, health status, income, wealth and retirement expectations (Olson, 1999; Haider and Solon, 2000; Kapteyn, Michaud, Smith and van Soest, 2006). In order to mitigate biases resulting from selection on observables, we construct inverse probability weights (Wooldridge, 2003). Kapteyn et al. (2006) show that using adjusted HRS weights based only on race, gender, household composition and birth cohort does not completely remove selection on observable biases because there are other observable characteristics that explain the decision to grant consent. As such, we follow their methodology and estimate a logit model at each wave utilizing demographic, health, work, and financial characteristics to predict the likelihood that a we have a matched SS record:  $\hat{p}(match_i = 1 | x_i)$ , where  $match$  is a dichotomous variable equal to one if we have matched SS data through the wave date and  $x$  represents the vector of aforementioned observable characteristics. We estimate this model separately for men and women and by cohort (using birth cohort for respondent level weights and entry cohorts for household level weights). The inverse probability weight is then simply  $(1 / \hat{p}(match_i = 1 | x_i))$ , which forms the non-response adjustment factor. Note that we construct weights for people with complete earnings records up to the wave date.

### A1. Household Level Weights

In creating household weights, we define a household as matched when we have matched SS earnings records for at least one member of the household through the wave date.

For the 1992 wave, we define the population at risk to include households with non-zero HRS household weights where at least one member of the household gave an interview in 1992. Table III.H.1A. outlines the sample and percentage matched for 1992. We define the population at risk in 1998 to include households with non-zero HRS household weights where at least one member of the household gave an interview in 1998 and the household belongs to one of the following entry cohorts: HRS, CODA, or WB (see Table III.H.1B & Table III.H.1C). In Table III.H.1C., we tabulate the percentage of households matched by entry cohort: we have matched records for 88 percent of the HRS households, 68 percent of the CODA households and 74 percent of the WB households. We define the population at risk in 2004 to include households with non-zero HRS household weights where at least one member of the household gave an interview in 2004 and the household belongs to one of the following entry cohorts: HRS or EBB (see Table III.H.1D and Table III.H.1E).

---

<sup>5</sup> This explanation is reprinted from *Prospective Social Security Wealth Measures: Technical Documentation*. Prepared by Kandice Kapinos with Charlie Brown, Michael A. Nolte, Helena Stolyarova and David R. Weir. Version 4.0, October 2009, pages15-17.

*Table III.H.1A. SSA Earnings Matches for 1992 Wave (Number of Households)*

At least one in HH Interviewed in 1992	7,703	
At least one in HH is age eligible		
(Total at-risk sample)	7,580	
Matched	6,411	85.03%
Unmatched	1,135	14.97%

*Table III.H.1B SSA Earnings Matches for 1998 Wave (Number of Households)*

At least one in HH Interviewed in 1998	14,999	
At least one in HH is age eligible	13,833	
In HRS, CODA, & WB (total at-risk sample)	9,614	
Matched	7,986	83.07%
Unmatched	1,628	16.93%

*Table III.H.1C. SSA Earnings Matches for 1998 Wave (Number of Households), by entry cohort*

<b>Entry cohort</b>	<b>In at- risk sample</b>	<b>Matched %</b>	<b>Unmatched %</b>
HRS	6,407	5,663 88	744 12
CODA	1,633	1,135 68	498 32
WB	1,574	1,188 75	386 25
Total	9,614	7,986	1,628

*Table III.H.1D.SSA Earnings Matches for 2004 Wave (Number of Households)*

At least one in HH Interviewed in 2004	14,321	
At least one in HH is age eligible	12,902	
In HRS, CODA, & WB (total at-risk sample)	7,874	
Matched	6,417	81.5%
Unmatched	1,457	18.5%

Table III.H.1E. SSA Earnings Matches for 2004 Wave (Number of Households), by entry cohort

<u>Entry cohort</u>	<u>In at- risk sample</u>	<u>Matched</u> <u>%</u>	<u>Unmatched</u> <u>%</u>
HRS	5,719	5,130 90	589 10
EBB	2,155	1,287 60	868 40
Total	7,874	6,417	1,457

We estimate separate weighted logit models for married couples<sup>6</sup>, single men and single women and by entry cohort. We then multiply the household level weight (AWGTH, FWGTH, and JWGTH for 1992, 1998, and 2004, respectively) by the non-response adjustment factor described above:

$$AWGTH * (1 / \hat{p}(match_i = 1 | x_i)) \quad [k]$$

where AWGTH is replaced by the appropriate wave household level weight. Next, we trim [k] at the fifth and ninety-fifth percentiles. Finally, we adjust the trimmed weights to reproduce the weighted distribution for the target sample cross-classified by age (6 categories: <46, 45-49, 50-54, 55-59, 50-64, 65+), sex, and race/ethnicity (3 categories: Hispanic, non-Hispanic Black, other).

## A2. Respondent Level Weights

For the 1992 wave, we define the population at risk as age-eligible (non-zero HRS respondent weight) individuals who gave an interview in 1992. Table III.H.2A. outlines the sample and fraction match for 1992. We define the population at risk in 1998 as age- eligible individuals who gave an interview in 1998 and belonged to the following birth cohorts: 1924-1930, 1931-1941, and 1942-1947 (see Table III.H.2B & Table III.H.2C).

Table III.H.2A. SSA Earnings Matches for 1992 Wave (Number of Respondents)

Interviewed in 1992	12,652	
Who are age-eligible (Total at risk sample)	9,794	
Matched	7,921	80.9%
Unmatched	1,873	19.1%

<sup>6</sup> We categorize a household as a “married couple” if the respondents define their marital status as married, married with spouse absent or partnered.

*Table III.H.2B. SSA Earnings Matches for 1998 Wave (Number of Respondents)*

Interviewed in 1998	22,881	
Who are age-eligible	20,050	
Total at risk sample	15,066	
Matched	8,281	55%
Unmatched	6,875	46%

*Table III.H.2C. SSA Earnings Matches for 1998 Wave, by birth cohort*

	<b>In at-risk</b>	<b>Matched %</b>	<b>Unmatched %</b>
Birth cohort = 1924-1930	3,692	2,020	1,672
		55	45
Birth cohort = 1931- 1936	4,244	2,189	2,055
		52	48
Birth cohort = 1937- 1941	4,080	2,219	1,951
		52	48
Birth cohort = 1942-1947	3,050	1,943	1,107
		64	36
Total	15,066	8,281	6,875

We estimate separate weighted logit models for men and women by birth cohort and wave. We then multiply the person level weight (AWGTR and FWGTR for 1992 and 1998, respectively) by the non-response adjustment factor described above:

$$AWGTR * (1 / \hat{p}(match_i = 1 | x_i)) \quad [I]$$

where AWGTR is replaced by the appropriate wave respondent level weight. Next, we trim [I] at the fifth and ninety-fifth percentiles. Finally, we adjust the trimmed weights to reproduce the weighted distribution for the target sample cross-classified by age (6 categories: <46, 45-49, 50-54, 55-59, 50-64, 65+), sex, and race/ethnicity (3 categories: Hispanic, non-Hispanic Black, other).

## Appendix B: Installation (Traditional License Only)

### B1. Distribution Set

The *Social Security Weights* data product is packaged for distribution in a .ZIP file, `sswgts.zip`. From this file extract the data file, the program statement file that matches your analysis environment, the data description (this file), and the codebook file. If you require a special file format or experience system problems, please contact the [HRS Help Desk](#). If all files are decompressed, they will require approximately 23MB of free space on your storage device.

#### *B1-1. Windows Environment*

Copy the ZIP file to the Windows folder where you plan to do your work. Do not use the built-in Windows decompression utility; use a third-party tool such as **7-ZIP** or **WinZIP** to extract the ZIP folder contents. When you are prompted for the pass-phrase, respond with the character string that you received via e-mail. The output will be the files listed in Table B1.

#### *B1-2. UNIX Environment*

Copy the ZIP file to the folder where you plan to do your work. Use the ZIP file decompression software installed on your system, (e.g. **7-zip**, **gunzip**) to decrypt and extract the ZIP folder contents. The output will be the files listed in Table B1.

#### *B1-3. Macintosh OS X Environment (10.4.x and above)*

Copy the ZIP file to the folder where you plan to do your work. Use **Stuffit-Expander** to decrypt and extract the ZIP folder contents. The output will be the files listed in Table B1.

### B1. Program Statements

Each data file comes with associated SPSS, SAS or Stata program statements to read the data. Files containing SPSS statements are named with an .sps extension, those with SAS statements with an .sas extension, and those with Stata statements with .do and .dct extensions.

#### *B2-1. Using the Files with SAS*

To create a SAS system file from an ASCII data (.da) file, load the .sas program statement files into the SAS Program Editor and reference the appropriate .da data files. If the \*.sas file is located in "c:\sswgts\sas" and the data file is located in "c:\sswgts\data", you can run the file as is. A SAS system file (\*.sas7bdat, assuming that you are using SAS V8.0 or higher) will be saved to directory "c:\sswgts\sas". If the files are not located in the specified directories, you will need to edit the \*.SAS file to reflect the proper path names prior to submitting the program statements.

#### *B2-2. Using the Files with SPSS*

To create an SPSS system file from an ASCII data (.da) file, load the .sps program statement files into the SPSS syntax editor window, reference the appropriate .da data files, and select the *Run>All* option. If the \*.sps file is located in "c:\sswgts\spss" and the data file is located in "c:\sswgts\data", you can run the file as is. An SPSS system file (\*.sav) will be saved to directory "c:\sswgts\spss". If the files are not located in the specified directories, you will need to edit the \*.sps file to reflect the proper path names prior to submitting the program statements.

#### *B2-3. Using the Files with Stata*

To create a Stata system file from an ASCII data (.da) file, three file types must be present: a .dct files, a .do files, and the ASCII data file containing the raw data for Stata to read. Files with the suffix ".dct" are Stata dictionaries used by Stata to describe the data. Files with the suffix ".do" are short Stata programs ("do files") which you may use to read in the data. Load the .do file into Stata and then submit it. If the SSWgtsDDv31.docx

\*.do and .dct files are located in "c:\sswgts\stata" and the data file is located in "c:\sswgts\data", you can run the .do file as is. If the files are not located in these directories, you must edit the \*.do and \*.dct files to reflect the proper path names before you run the files.

*Table B1: Contents of the Distribution Package*

<b>Directory<sup>7</sup></b>	<b>File</b>	<b>Type</b>
c:\sswgts\data\	SSWG TSA_R.da SSWG T SF_R.da SSWG T SK_R.da	Data files (ASCII text)
c:\sswgts\docs\	SSWG TS.txt SSWG T SDD.pdf	Codebook file (ASCII text) Data Description (this document)
c:\sswgts\sas\	SSWG TSA_R.sas SSWG T SF_R.sas SSWG T SK_R.sas	SAS program statements
c:\sswgts\spss\	SSWG TSA_R.sps SSWG T SF_R.sps SSWG T SK_R.sps	SPSS program statements
c:\sswgts\stata\	SSWG TSA_R.dct/SSWG TSA_R.do SSWG T SF_R.dct/SSWG T SF_R.do SSWG T SK_R.dct/SSWG T SK_R.do	Stata dictionary/"do" files

---

<sup>7</sup> When using HRS data products, you should feel free to create the directory structure that is most suitable for your needs. By using the suggested directory structure (or a Unix equivalent), you will not have to change the path name references in the data descriptor files. If you want to use a different structure, just change the directory references in the program files.