

Health and Retirement Study
Prospective Social Security Wealth Measures of Pre-Retirees
Wave 10

Data Description and Usage

Chichun Fang and Kandice Kapinos
Institute for Social Research
University of Michigan

Version 1.0
March 2016

Contents

1	Overview	1
2	Background	1
3	Benefits Calculation	2
3.1	Earnings Projection and Benefits Calculation for the Matched Sample	2
3.2	Benefits Imputation for the Unmatched Sample	4
3.2.1	Respondent Imputations	5
3.2.2	Spouse Imputations	7
4	Wealth Calculation	7
4.1	Types of Benefits	7
4.2	Wealth Based on Benefits from Respondent's Own Earnings History	8
4.3	Wealth Based on Spousal and Survivor Benefits	9
4.4	Total Household Wealth	10
4.5	Rounding and Top-Coding	10
4.6	Wealth for Respondents Who Have Claimed	11
5	Distribution Information	11
5.1	Variable Naming Conventions	12
5.2	Identification Variables	13
5.2.1	HHID: Household Identifier	13
5.2.2	PN: Person Number	13
5.3	Program Statements	13
5.3.1	Using the Files with SAS	13
5.3.2	Using the Files with SPSS	14
5.3.3	Using the Files with Stata	15
5.4	Registration and Downloading the Data	15
5.4.1	Conditions of Use	15
5.4.2	Publications Based on Data	16
5.5	If You Need to Know More	16
5.5.1	HRS Website	17
5.5.2	Contact Information	17
	References	18

1 Overview

In this document, we outline assumptions, calculations, and other details pertaining to the construction of Social Security wealth measures for respondents in the Health and Retirement Study (HRS), a national longitudinal study of the economic, health, marital, family status and public and private support systems of older Americans. The National Institute on Aging provided funding (NIA U01 AG009740), with supplemental support from the Social Security Administration (SSA). The study is conducted by the Institute for Social Research at the University of Michigan.

This release of the *Prospective Social Security Wealth Measures of Pre-Retirees* data set consists of respondent-level benefits wealth estimates constructed from the employment section of the 2010 HRS (wave 10) and the restricted SSA summary and detailed earnings and benefits files. In this public use file, we calculate wealth only for individuals who have not yet retired. Each individual is uniquely identified by the concatenation of the household ID and the person number, HHID and PN. We organize the data to match the organization of the RAND HRS data files.

By receiving the data, which has been freely provided, you agree to use them for research and statistical purposes only and make no effort to identify respondents. In addition, you agree to send HRS a copy of any publications you produce based on the data. See Sections [5.4.1](#) and [5.4.2](#) for additional details.

2 Background

We utilize information from HRS respondents who authorized the University of Michigan’s Institute for Social Research to obtain administrative records from the Social Security Administration to calculate and impute Social Security benefit wealth to respondents in the 2010 wave of the HRS. In the following, we call the HRS respondents whom we have administrative records for the “matched sample” and those whom we do not the “unmatched sample.”

There are two major components of administrative records that we have for the matched sample: the earnings records and the benefits records. The earnings records contain the amounts of yearly earnings covered by Social Security between as early as 1951 and the most recent year we have a

valid consent.¹ For the respondents who have started claiming benefits, the benefit records contain type of benefit (disability, retirement, spousal, survival, etc.) and monthly benefit amounts. In this data set, we calculate Social Security benefits wealth only for those in the matched sample who have not claimed retirement benefit yet.

Sections 3 covers the procedures and assumptions used to calculate or impute monthly Social Security benefits, which are used to calculate various components of Social Security benefit wealth detailed in Section 4. For the unmatched sample, we rely on the self-reported information in the survey to determine whether a respondent has already claimed the old-age benefit. Again we retain only those who have not claimed their benefits yet. We impute their Social Security benefit using the procedures detailed in Section 3.2.

3 Benefits Calculation

3.1 Earnings Projection and Benefits Calculation for the Matched Sample

We use the AnyPIA program (Office of the Actuary, v2015.1) to compute each respondent’s average indexed monthly earnings (AIME) and primary insurance amount (PIA). For each respondent, we calculate AIME, PIA, and benefits amount at three claim age: age 62 (early retirement), full retirement age, and age 70 (delayed retirement). We discuss the program’s algorithm and our relevant assumptions in this section.

The AnyPIA program determines eligibility and coverage based on the individual’s earnings record. We have earnings as early as 1951. We project earnings beyond 2010 following Mitchell, Olson and Steinmeier (2000) for individuals who have yet to attain the relevant claim age (62, full retirement age, or 70) in 2010. Earnings in the five-year period preceding the wave year (t) are indexed to the $t - 1$ levels using the National Average Wage Index (AWI) of the SSA based on the Trustees Report for each wave year. The indexed wages are then averaged, with years $t - 1$ through $t - 5$ single year earnings given weights 5, 4, 3, 2, and 1, respectively. Therefore, the projected real

¹Before 2004, HRS respondents had to re-consent each time they are surveyed, as the consent is only valid through the survey year. Starting from 2006, the HRS uses prospective consents, which, once given, remain valid unless being revoked by the respondent. Due to the different types/timing of consent given, the end year of earnings records differs by respondents.

earnings for the wave date (t) are:

$$Y_t = \left(1 + CPI_t\right) \left(5 \cdot Y_{t-1} + 4 \cdot Y_{t-2} \cdot \frac{AWI_{t-1}}{AWI_{t-2}} + 3 \cdot Y_{t-3} \cdot \frac{AWI_{t-1}}{AWI_{t-3}} + 2 \cdot Y_{t-4} \cdot \frac{AWI_{t-1}}{AWI_{t-4}} + 1 \cdot Y_{t-5} \cdot \frac{AWI_{t-1}}{AWI_{t-5}}\right) \quad (1)$$

where Y_t represents the SS earnings in year t and $\frac{AWI_{t-1}}{AWI_{t-i}}$ is the index factor that inflates earnings in year $t - i$ to year $t - 1$ dollars. Then, AnyPIA automatically indexes Y_t upward each year until the corresponding year, using the National AWI from the wave date Trustee Report.

For most of the respondents in this data set, year $t - 1$ is 2009. For the respondents who we do not have the most update earnings history, year $t - 1$ is the last year that the respondent's earnings records is available to us. Regardless what $t - 1$ is and how old the respondent was in year $t - 1$, We project the earnings history up to the year before the relevant claim age. That is, we project earnings history up to the year before age 62 to calculate early retirement benefits, up to the year before full retirement age to calculate full retirement benefits, and up to the year before age 70 to calculate delayed retirement benefits.

In cases where a respondent has \$0 earnings in year $t - 1$, we examine his self-reported employment status in year t (2010 for most of the respondents). If the respondent reported working for pay in year t , we use the weighted average method as just described. However, if the respondent reported not working for pay in year t , we project earnings of \$0 beyond t . In other words, if a respondent was working in year t , we do not allow the respondent to quit the labor market before the relevant claim age.. If a respondent was not working in year t , we do not allow the respondent to re-enter the labor market.

Given a respondent's birth date, the year and month that corresponds to his or her relevant claim age (62, full retirement age, or 70) can be calculated. For each respondent and each claim age, the relevant claim date and month as well as the corresponding earnings history are entered into AnyPIA to calculate AIME, PIA, and retirement benefits at that claim age.

AIME is computed as the individual's highest indexed earnings averaged over the appropriate number of "computation years" (see §703.1 of the SSA Handbook). The PIA is then a function of the AIME and "bend points", which are also indexed over time and are designed to form a

regressive relationship between AIME and PIA. Finally, monthly benefits are determined by PIA and retirement age. Benefits are reduced/increased if the respondent retires before/after the full retirement age, which depends on birth year and is between 66 and 67 for most of the sample in this data set. Depending on the timing of retirement relative to full retirement age, the monthly retirement benefit can be as low as 70% of or as high as 132% of PIA at claim age.

In our benefits calculation, we exclude earnings from non-covered (i.e. not covered by Social Security) jobs as well as earnings not subject to the payroll tax. Hence, we use AnyPIA to calculate benefits based on earnings from covered jobs only. For the respondents who have accumulated pension benefits in non-covered jobs, their Social Security benefits are subject to reduction according to the Windfall Elimination Provision (WEP) or Government Pension Offset (GPO) (see §718 and §1836 of the SSA Handbook for more details).² In this data set, we are not able to do the WEP and/or GPO benefits adjustment because we do not have enough information about the pension benefits in non-covered jobs.

Once we obtain early, full, and delayed monthly retirement benefits for each of the matched sample, we adjust these numbers to 2010 U.S. dollars based on the intermediate economic assumption in the 2010 version of the *Annual Report of the Board of Trustees of the OASDI Trust Funds*, summarized in Table 1. The same economic assumptions are also used to calculate the Social Security benefits wealth in Section 4.

TABLE 1: Economic Assumptions

Wave Year	Nominal Interest	Inflation	Real Interest
2010	5.8 %	2.8%	2.9%

3.2 Benefits Imputation for the Unmatched Sample

We impute own and spouse PIA only in cases where the respondent and/or the respondent’s spouse are unmatched to Social Security earnings records. In cases where we have some (but not all) earnings records for a respondent, we extrapolate the earnings history based on equation (1)

²Gustman, Steinmeier, and Tabatabai (2014) estimates that about 3.5% of retired HRS respondents are subject to WEP or GPO.

and rely on AnyPIA to calculate PIA for this respondent. For example, if a respondent only has earnings records as recent as 2004, this respondent still is counted toward the matched sample and the earnings from 2005 and beyond are extrapolated.

We have created two variables to indicate our imputation strategy for respondent and spouse PIAs, respectively: R10IMPUTE and S10IMPUTE. We define these indicator variables and the corresponding imputation strategies in the following. While the imputed wealth is included in the public release of this dataset, the variables R10IMPUTE and S10IMPUTE are only available as restricted data.

3.2.1 Respondent Imputations

The indicator variable for respondent imputations, R10IMPUTE, takes on the following values:

- 1, if the respondent belongs to the matched sample (no imputation);
- 2, if the respondent is unmatched and expects to claim Social Security benefits (self-reported);
- 3, if the respondent is unmatched and does not expect to claim Social Security benefits or does not report expectations according to the HRS;
- Missing, if the respondent is already receiving retirement benefits according to the Master Beneficiary Records (the matched sample) or the HRS survey (the unmatched sample).

In cases where R10IMPUTE is 2 or 3, we impute the respondent’s own PIA using the nearest neighborhood matching method. Starting with the sample of matched respondents, we regress the logarithm of respondent’s own PIA at the full retirement age (calculated in the previous section) on several demographic and labor force characteristics as of 2010.³ To allow the coefficients in the

³We include the following explanatory variables: respondent’s age, three marital status indicators (married, divorced, or widowed), eight region of residence dummy variables, two dummy variables for minority (non-Hispanic black or Hispanic), two education indicators (some college or college), an indicator for whether the respondent was born outside of the U.S., two labor force status indicators (working full-time or part-time), a self-employment indicator, years of work experience (a linear term but capped at 35), a union status indicator, three pension plan indicators (having a defined benefit plan, having a defined contribution plan, or having both), a health insurance indicator, a home ownership indicator, logarithm of annual wage for the current or last job, logarithm of total household income, occupation and industry indicators for the current or last job, and the number of years in federal jobs prior to 1984. The number of years in federal jobs before 1984 is included because federal employees were not covered by Social Security before 1984. For the logarithm of wage and income variables, we use the inverse hyperbolic sine transformation, $\ln(y + \sqrt{y^2 + 1})$, rather than $\ln(y)$ so the observations with zero earnings or income are not dropped.

imputation equation to differ by gender and to include as many lagged terms of annual wages as possible, we split the matched sample into four subsamples: (a) males whose first HRS interview occurred after (not including) 2004; (b) females first interviewed after 2004; (c) males first interviewed in or before 2004; and (d) females first interviewed in or before 2004. For subsamples (a) and (b), we use the right hand side variables as specified in the footnote in the imputation equation. For subsamples (c) and (d), we use the same set of variables plus the logarithm of annual earnings in 2008, 2006, and 2004.⁴

The unmatched sample is split into four subsamples using the same categorization. We then use the coefficients from the corresponding imputation equation to “predict” PIA for the whole (matched and unmatched) sample and sort all observations based on their predicted PIA.⁵ For each member of the unmatched sample, we find the matched respondent who has the closest predicted PIA and assign the actual PIAs at various ages (62, FRA, and 70) of this matched respondent as the imputed PIAs for this unmatched respondent. If an unmatched respondents have more than one such “nearest neighbors” (for example, when the unmatched respondent has a predicted PIA sandwiched between predicted PIAs from two matched respondents in equal distance), one of the nearest neighbors is randomly chosen.

Each respondent in the unmatched sample is assigned the imputed PIAs based on the nearest matching. The imputed PIAs for the unmatched sample are used along with the actual PIAs from the matched sample to calculate Social Security benefits wealth. To facilitate the calculation of wealth at the household level (see Section 4 for more details), we also construct an indicator variable for benefits imputation of the spouse, which is introduced in the very next section.

⁴We use a sequence Monte Carlo experiments based on the matched sample only to test the robustness of this imputation strategy. More specifically, we use the actual PIAs from (randomly-chosen) half of the matched sample to “impute” the PIAs for the other half and compute the difference between the imputed and actual PIAs. The results from our Monte Carlo exercise suggests that including earnings before 2004 for respondents interviewed before 2004 actually increases the mean squared error of the imputed PIA. Using a two-step model that simultaneously imputes Social Security coverage and PIA also increases the magnitude of error.

⁵These predicted PIAs are estimated with subsample-specific specifications and coefficients, but the predicted PIAs are sorted regardless of subsample.

3.2.2 Spouse Imputations

The indicator variable for spouse imputations, S10IMPUTE, is defined very similarly as R10IMPUTE. After each respondent in the unmatched sample is assigned the imputed PIAs, we aggregate the information to the household level. Note that, in a household that consists of a married couple, the first member's R10IMPUTE is the second member's S10IMPUTE, and the the first member's S10IMPUTE is the second member's R10IMPUTE. The only difference between R10IMPUTE and S10IMPUTE is that S10IMPUTE is set to missing for non-married respondents.

4 Wealth Calculation

4.1 Types of Benefits

We focus on three distinct types of Social Security benefits in this dataset. A retired worker age 62 or over who is fully insured is eligible to receive monthly retirement benefits (also called old-age insurance benefits) that are based on his or her own earnings history (see §300 of the SSA Handbook). In addition, some individuals may be eligible for other auxiliary benefits, including spousal and survivor benefits, which are based on the spouse's earnings history (see §305 of the SSA Handbook). In this case, the fully insured retired worker receives a monthly benefit that is comprised of both the retirement benefits and spousal (when both the respondent and the spouse are alive) or survivor (when the respondent is widowed) benefits. We offer the following examples borrowed from Panis et al. (2000) to explain how the auxiliary benefits are calculated.

Example 1 Mary and Paul are married. Mary gets \$200 a month in SSA retirement based on her own work history. Paul gets \$1,000 a month for his SSA retirement insurance benefits. Mary is entitled to an auxiliary spouse benefit in the amount of \$500 from Paul's account. Since the \$500 is greater than her own retirement insurance benefit amount (\$200), she is dually entitled. She will receive \$200 against her own account and an additional \$300 against Paul's. We refer to the \$200 portion of this benefit as her retirement insurance benefit and the \$300 portion as her incremental auxiliary spouse benefit. Note that she actually receives this all in one check in the amount of \$500. If Mary's own benefit were \$600 instead of \$200, she would not be termed dually entitled and would

have no auxiliary spouse benefit. Similarly, Paul is not dually entitled on Mary’s account because he is only eligible for half her benefit (\$100), which is less than what he receives on his own account. Therefore, for the construction of the wealth measures in this dataset, here is a recapitulation of how we define the types of benefits in this case:

TABLE 2

	Own Benefit	Spousal Benefit
Mary	\$200	\$300
Paul	\$1,000	\$0

Example 2 Suppose Paul passes away. As Paul’s widow, Mary is entitled to the full amount of his entitlement, \$1000. Mary would receive \$200 based on her own lifetime earnings and \$800 from Paul’s account, for a total monthly benefit in the amount of \$1000. We refer to the \$200 portion again as her “retirement insurance benefit” and the \$800 as her “incremental survivor benefit.” Now, the data would change:

TABLE 3

	Own Benefit	Spousal Benefit	Survivor Benefit
Mary	\$200	\$0	\$800
Paul	\$0	\$0	\$0

In our calculation, we separate a respondent’s benefits into three components: (1) retirement benefits based on the respondent’s own earnings history, (2) spousal benefit when both couples are alive, and (3) the survivor benefit when the respondent is widowed. By definition, only the first component is relevant for non-married respondents. We calculate the present value of these three separate streams of income adjusted by the appropriate survival probabilities.

4.2 Wealth Based on Benefits from Respondent’s Own Earnings History

Using the SSA mortality tables⁶ and the relevant interest and inflation rates as projected from the *Trustee Report*, we calculate the present value of the stream of Social Security retirement insurance

⁶We use expected survival probabilities up to age 119 based on the 2010 version of gender-specific SSA cohort mortality tables.

benefits based on the retirement insurance benefit for each claim date. The expected discounted present value of benefits is given by:

$$PV_{T_0} = PIA_{T_0} \sum_{t=T_0}^{119} {}_tP_{2010} \cdot (1+r)^{-(t-T_0)} \quad (2)$$

where PIA_{T_0} is the annual benefit payment at the projected claim date ($T_0=62$, FRA, or 70) after early/delayed retirement adjustment calculated or imputed in SectionBenefit, ${}_tP_{2010}$ is the probability of being alive at age t conditional on having survived through 2010, and r is the nominal interest rate.

We sum discounted benefits from the claim year ($t = T_0$) until the age of 119. The first term represents the retirement benefits conditional on claim age c . We adjust this value by the probability of survival through age t and discount it back to claim age (T_0) U.S. dollars. We use yearly compounding since our mortality data are at yearly intervals. To make these values comparable across individuals, we then move PV_{T_0} from claim year dollars to 2010 U.S. dollars, which is what is reported in the data files. Also, note that we adjust the first year of benefits to account for the fractional consumption of benefits in that year. For example, if an a respondent starts receiving benefit in July, then we use one-half of the relevant PIA_{T_0} for the first year in the summation. We assign a missing value for wealth in cases where the individual is older than a particular claim date as of the wave date. For example, a respondent who is age 68 as of 2010 will have missing values for R10SSWRER (wealth based on own earnings record and early retirement claim age) and R10SSWRNR (wealth based on own earnings record and full retirement claim age).

4.3 Wealth Based on Spousal and Survivor Benefits

In order to calculate the auxiliary spouse benefits for a married respondent, we need the spouse's retirement benefits (based on the spouse's own earnings history). The calculation of spousal benefits is complicated by various possible combinations of claiming ages between the couple. In this data set, we assume that both couple claim at the same age to simply the calculation.⁷

⁷We also assume that, if a couple is married in 2010, they stay married though the remaining of their lives. Furthermore, to be eligible for spousal benefits, the marriage needs to last at least 10 years. We use the length of marriage at the relevant claim age, rather than as of 2010, to determine such eligibility.

Let i and j represent the respondent and the spouse, respectively. Respondent i 's total Social Security benefits wealth, including own retirement benefits and spousal benefits, is defined as (assuming adjustments for early/late claiming have already been made to PIAs):

$$\begin{aligned}
SSWealth_i &= PIA_i \sum_{t=T_0}^{119} {}_tP_{2010,i} (1+r)^{-(t-T_0)} + \\
&\quad \max(0, 0.5 \cdot PIA_j - PIA_i) \sum_{t=T_0}^{119} {}_tP_{2010,i} \cdot {}_tP_{2010,j} (1+r)^{-(t-T_0)} + \\
&\quad \max(0, PIA_j - PIA_i) \sum_{t=T_0}^{119} {}_tP_{2010,i} \cdot (1 - {}_tP_{2010,j}) (1+r)^{-(t-T_0)}
\end{aligned} \tag{3}$$

where PIA_j is the annual retirement benefits amount for spouse j at the projected claim date, ${}_tP_{2010,j}$ is the probability spouse j is alive at age t conditional on having survived through 2010, and r is the nominal interest rate.

The first term represents social security benefits wealth based on respondent i 's own earnings history and is the same as PV_{T_0} in equation (2). The second term represents wealth based on the incremental auxiliary spouse benefit, which is adjusted by the likelihood that both spouses are alive each period; the third term represents wealth based on the incremental survivor (widow or widower) benefit, also adjusted survival probabilities. Note that incremental auxiliary spouse and survivor wealth are zero for the higher earner (by definition). The wealth estimates are reported in 2010 dollars.

4.4 Total Household Wealth

To make these values comparable across individuals, household wealth is reported in 2010 U.S. dollars. Hence, for members i and j in the same household, total household wealth is defined as:

$$HHSSWealth_{ij} = SSWealth_i + SSWealth_j \tag{4}$$

4.5 Rounding and Top-Coding

In equation (3), the wealth based on the respondent's own earnings history (the first term on the right hand side) and the wealth based on spousal and survival benefits (the sum of the second and third terms on the right hand side) are both rounded to the nearest \$100 and then top-coded at

the top 2 percent.⁸ $SSWealth_i$ is then calculated based on these rounded and top-coded terms. Neither $SSWealth_i$ in equation (3) nor $HHSSWealth_{ij}$ in equation (4) is further rounded and/or top-coded.

4.6 Wealth for Respondents Who Have Claimed

Individuals are designated as already having claimed if they have started receiving benefits as of the relevant wave date according to the Social Security Master Beneficiary Record file (for the matched sample) or the HRS survey (for the unmatched sample). In this version of the data, we flag these individuals using R10CLAIMED and S10CLAIMED. They will have missing values for all the SS wealth prediction variables.

5 Distribution Information

The Prospective Social Security Wealth Measures of Pre-Retirees data are distributed in 1 data file at the respondent level with Social Security benefit wealth estimates in 2010. The records in the data files are sorted in order by these primary identifiers. The data are provided in ASCII format, with fixed-length records. Use associated SAS, SPSS, or Stata program statements to read the data into the analysis package of your choice. The file is packaged for download from our website in ZIP format.

The following extensions are used for the six different types of distribution files:

- *.da for data files,
- *.sas for SAS program statements,
- *.sps for SPSS program statements,
- *.do for Stata do statements,
- *.dct for Stata dictionary statements, and
- *.txt for codebook files.

⁸For the purpose of top-coding, “top 2%” of each variable is defined as top 2% of the distribution excluding missing values and zero.

For example,

- `SSWEALTHP2010.da` contains ASCII data.
- `SSWEALTHP2010.sas` contains corresponding SAS program statements,
- `SSWEALTHP2010.sps` contains corresponding SPSS program statements,
- `SSWEALTHP2010.do` contains corresponding Stata do statements,
- `SSWEALTHP2010.dct` contains corresponding Stata dictionary statements, and
- `SSWEALTHP2010.txt` contains the ASCII codebook.

The Social Security benefit wealth estimates data are provided in ASCII format, with fixed-length records. Use the associated SAS, SPSS, or Stata program statements to read the data into the analysis package of your choice. In addition, you will probably want to download the codebook file (`SSWEALTHW10.txt`) and the data description (this document).

5.1 Variable Naming Conventions

The variable names are consistent with the pattern used in the RAND HRS data. The first character indicates whether the variable refers to the respondent, “R”; the spouse, “S”; or the household, “H”. The second and third character corresponds to the wave to which the variable pertains: “10” for 2010. All Social Security wealth measures are then followed by “SSW”. The seventh character denotes the source of the benefit: “R” for retirement insurance benefit. The eighth and ninth characters denote the age at claim: “ER” for claiming at the early retirement age, “NR” for claiming at the normal or full retirement age, and “XA” for claiming at age 70.

We calculate benefits assuming claim ages that the respondent has not yet attained as of 2010, otherwise, we assign missing values if the respondent is older than a particular claim age. We also assign missing values if the respondent has already started claiming benefits as evidenced by being in the Social Security Master Beneficiary Records file. Table 4 shows which wealth values we calculate for individuals of a given age in 2010.

Table 5 lists the variables included in the data file and some brief descriptions.

TABLE 4

Age in 2010	Assuming Claim Age of:		
	Early (Age 62)	Full	Late (Age 70)
70 and Above	–	–	–
FRA-69	–	–	X
63-FRA	–	X	X
62 or Younger	X	X	X

5.2 Identification Variables

5.2.1 HHID: Household Identifier

In the initial wave of data collection, each sample household is assigned a six digit Household Identifier, HHID. This variable is stable across waves of data collection and uniquely identifies the original household and any households derived from that household in subsequent waves of data collection.

5.2.2 PN: Person Number

In combination with HHID, PN uniquely identifies a respondent or respondent’s spouse or partner. PNs are unique within an original household (HHID). The PN assigned to a particular respondent does not change across waves. PN has three digits.

5.3 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 `*.spss`, those with SAS statements with `*.sas` extension, and those with Stata statements with `*.do` and `*.dct` extensions.

5.3.1 Using the Files with SAS

To create a SAS system file for a particular dataset, two file types must be present for that dataset: `*.sas` program statement files and `*.da` data files. To create a SAS system file, load the `*.sas` file into the SAS Program Editor.

If the `*.sas` file is located in “`c:\sswealthp2010\sas`” and the data file is located in “`c:\sswealthp2010\my data`”, you can run the file as is. A SAS system file (`*.sas7bdat`) will be

TABLE 5: List of Variables

Variable Name	Brief Description
HHID	Household Identifier
PN	Person Number
R10SSWRER	Respondent's benefit wealth based on own earnings history, assuming claim age at 62; calculated as in equation (2).
R10SSWRNR	Respondent's benefit wealth based on own earnings history, assuming claim age at FRA; calculated as in equation (2).
R10SSWRXA	Respondent's benefit wealth based on own earnings history, assuming claim age at 70; calculated as in equation (2).
R10CLAIMED	An indicator that equals 1 if the respondent is already receiving benefits.
S10SSWRER	Respondent's benefit wealth based on spousal and survival benefits, assuming claim age at 62; sum of the second and third components in equation (3).
S10SSWRNR	Respondent's benefit wealth based on spousal and survival benefits, assuming claim age at FRA; sum of the second and third components in equation (3).
S10SSWRXA	Respondent's benefit wealth based on spousal and survival benefits, assuming claim age at 70; sum of the second and third components in equation (3).
S10CLAIMED	An indicator that equals 1 if the respondent's spouse is already receiving benefits.
H10SSWRER	Total household benefit wealth, assuming claim age at 62; calculated as in equation (4).
H10SSWRNR	Total household benefit wealth, assuming claim age at FRA; calculated as in equation (4).
H10SSWRXA	Total household benefit wealth, assuming claim age at 70; calculated as in equation (4).

saved to directory “c:\sswealthp2010\sas”. If the files are not located in the specified directories, you will need to edit the *.sas file to reflect the proper path names prior to running the file.

5.3.2 Using the Files with SPSS

To create an SPSS system file for a particular dataset, two file types must be present for that dataset: *.sps program statement files and *.da data files. To create an SPSS system file, open the *.sps file in SPSS as an SPSS Syntax File.

If the *.sps file is located in “c:\sswealthp2010\spss” and the data file is located in “c:\sswealthp2010\DATA”, you can run the file as is. An SPSS system file (*.sav) will be saved to directory “c:\sswealthp2010\spss”. If the files are not located in the specified directories, you will need to edit the *.sps file to reflect the proper path names prior to running the file.

5.3.3 Using the Files with Stata

To use Stata with a particular dataset, the following three file types must be present for that dataset: *.dct files, *.do files, and *.da data files.

Files with the suffix *.da contain the raw data for Stata to read. Files with the suffix *.dct are Stata dictionaries used by Stata to describe the data. Files with the suffix *.do are short Stata programs (“do files”) which you may use to read in the data. Load the *.do file into Stata and then submit it.

If the *.do and *.dct files are located in “c:\sswealthp2010\Stata” and the data file is located in “c:\sswealthp2010\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. Note that the variable names provided in the *.dct files are uppercase. If you prefer lower case variable names, you may wish to convert the *.dct files to lower case prior to use.

5.4 Registration and Downloading the Data

HRS data are available free to researchers and analysts at the HRS website. In order to obtain public release data, you must first register at our website. Once you have completed the registration process, your username and password will be sent to you via email. Your username and password are required to download any data files. By registering all users, we are able to document for our sponsors the size and diversity of our user community allowing us to continue to collect these important data. Registered users receive user support, information related to errors in the data, future releases, workshops, and publication lists. The information you provide will not be used for any commercial use and will not be redistributed to third parties.

5.4.1 Conditions of Use

By registering, you agree to the Conditions of Use governing access to Health and Retirement Study public release data. You must agree:

- Not to attempt to identify respondents;

- Not to transfer data to third parties except as specified;
- Not to share your username and password;
- To include specified citations in work based on HRS data;
- To provide information to us about publications based on HRS data;
- To report apparent errors in the HRS data and documentation files;
- To notify us (via our website) of changes in your contact information.

5.4.2 Publications Based on Data

As part of the data registration process, you agree to include specified citations and to inform HRS of any papers, publications, or presentations based on HRS data. Please send a copy of any such publications in PDF format via e-mail to hqsquestions@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
Institute for Social Research
University of Michigan
426 Thompson Street
Ann Arbor, MI 48104

5.5 If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the Social Security Wealth (Public) data. If you have questions or concerns that are not adequately covered here or on our website, or if you have any comments, please contact us. We will do our best to provide answers.

5.5.1 HRS Website

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 browser to the HRS website: <http://hrsonline.isr.umich.edu>

5.5.2 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 website
- E-mail: hqsquestions@umich.edu

- Postal Service:

Health and Retirement Study

Institute for Social Research

University of Michigan

426 Thompson Street

Ann Arbor, MI 48104

References

- Gustman, Alan L., Thomas L. Steinmeier, and Nahid Tabatabai.** 2014. “The Social Security Windfall Elimination and Government Pension Offset Provisions for Public Employees in the Health and Retirement Study.” *Social Security Bulletin*, 74(3): 55–69.
- Mitchell, Olivia S., Jan Olson, and Thomas L. Steinmeier.** 2000. “Social Security Earnings and Projected Benefits.” In *Forecasting Retirement Needs and Retirement Wealth.* , ed. Olivia S. Mitchell, P. Brett Hammond and Anna M. Rappaport, Chapter 13. University of Pennsylvania Press.
- Panis, Constantijn, Ronald Euler, Cynthia Grant, Melissa Bradley, Christin E. Peterson, Randall Hirscher, and Paul Steinberg.** 2000. *SSA Program Data User’s Manual*. RAND.