

Health and Retirement Study  
Prospective Social Security Wealth Measures of Pre-Retirees  
Waves 10 and 13  
Data Description and Usage

Chichun Fang  
Institute for Social Research  
University of Michigan

Version 2  
October 2021

---

\*This documentation is partially built upon the work of Kandice Kapinos.

# Contents

<b>1</b>	<b>Overview</b>	<b>1</b>
<b>2</b>	<b>Background</b>	<b>1</b>
<b>3</b>	<b>Benefits Calculation</b>	<b>2</b>
3.1	Earnings Projection and Benefits Calculation for the Matched Sample . . . . .	2
3.2	Benefits Imputation for the Unmatched Sample . . . . .	5
3.2.1	Respondent Imputations . . . . .	5
3.2.2	Spouse Imputations . . . . .	7
<b>4</b>	<b>Wealth Calculation</b>	<b>7</b>
4.1	Types of Benefits . . . . .	7
4.2	Wealth Based on Benefits from Respondent's Own Earnings History . . . . .	9
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
<b>5</b>	<b>Distribution Information</b>	<b>11</b>
5.1	Variable Naming Conventions . . . . .	12
5.2	Identification Variables . . . . .	14
5.2.1	HHID: Household Identifier . . . . .	14
5.2.2	PN: Person Number . . . . .	14
5.3	Program Statements . . . . .	14
5.3.1	Using the Files with SAS . . . . .	14
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 . . . . .	16
5.4.2	Publications Based on Data . . . . .	16
5.5	If You Need to Know More . . . . .	17
5.5.1	HRS Website . . . . .	17
5.5.2	Contact Information . . . . .	17
	<b>References</b>	<b>18</b>

# 1 Overview

In this documentation, we outline assumptions, calculations, and other details pertaining to the construction of Social Security Benefit wealth estimates 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 Benefit Wealth Estimates and Imputations* data set consists of respondent-level Social Security Benefit wealth estimates constructed from the employment section of the 2010 and 2016 HRS (waves 10 and 13) and the restricted SSA summary and detailed earnings and benefits files. In this public use file, we calculate wealth only for respondents who have not yet claimed old-age benefits. Each respondent is uniquely identified by the concatenation of the household ID and the person number, HHID and PN.

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 as of the 2010 and 2016 waves of the HRS. In the following, we call the HRS respondents for whom we have administrative records the “matched sample” and those for whom we do not the “unmatched sample.”

There are two major components of administrative records that we have for the matched sample: earnings records and 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.<sup>1</sup> For the respondents who have claimed benefits, Master Beneficiary Records (MBR) contain information on the type(s) of benefits (disability, old-age, spousal, survival, etc.) and monthly benefits amounts. In this data set, we calculate Social Security Benefit wealth only for those in the matched sample who have not yet claimed their old-age benefits.

Sections 3 covers the procedures and assumptions used to calculate or impute monthly Social Security Benefit, 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 benefits. 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, v2016.1) to compute each respondent’s average indexed monthly earnings (AIME) and primary insurance amount (PIA). For each respondent, we calculate the AIME, the PIA, and the benefits amount at three claim ages: 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 records. We have earnings as early as 1951. Only earnings from jobs (including self-employment) covered by Social Security are used, and the term “earnings” refers to “earnings covered by Social Security” for the rest of this documentation. We project earnings beyond 2016 following Mitchell, Olson and Steinmeier (2000) for individuals who have yet to attain the corresponding claim ages (62, full retirement age [FRA], or 70) in 2016. Earnings in the five-year period preceding the wave year ( $t$ ) are indexed to  $t - 1$  levels using the national Average Wage Index (AWI). The indexed wages are then averaged, with years  $t - 1$  through  $t - 5$  single year earnings given weights 5, 4, 3,

---

<sup>1</sup>Prior to 2006, consents were only valid for that particular wave when they were given. Starting from 2006, the HRS uses prospective consents that are valid for 6 years for earnings records. Due to the different types/timing of consent given, the end year of earnings records differs by respondents.

2, and 1, respectively.

For most of the respondents in this data set, year  $t - 1$  is 2015. For the respondents for whom we do not have the most updated earnings history, year  $t - 1$  is the last year that the respondent’s earnings records are available to us. Regardless of the year associated with  $t - 1$  and how old the respondent was in year  $t - 1$ , we project the earnings through the year before the relevant claiming age. That is, we project earnings through the year before age 62 to calculate early retirement benefits, through the year before full retirement age to calculate full retirement benefits, and through the year before age 70 to calculate delayed retirement benefits.

The projected real earnings for the wave date ( $t$ ) can be expressed as:

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

where  $Y_t$  represents the 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.

In cases where a respondent has \$0 earnings in year  $t - 1$ , we examine their self-reported employment status in year  $t$  (2016 for most 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 assume the respondent does not quit the labor market before the relevant claim age. If a respondent was not working in year  $t$ , we assume the respondent does not re-enter the labor market.

Given a respondent’s birth date, the year/month that corresponds to his or her relevant claiming age (62, full retirement age, or 70) can be determined. For each respondent and each claim age, the relevant claim date (year/month) as well as the corresponding earnings history are entered into AnyPIA to calculate the AIME, the PIA, and the amount of old-age benefits at that claiming 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 the PIA and claim age. Benefits are reduced/increased if the respondent claims 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 the full retirement age, the monthly retirement benefits at the claim age can be as low as 70% or as high as 132% of the PIA.

For respondents who have accumulated pension benefits from jobs not covered by Social Security, their Social Security Benefit 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).<sup>2</sup> The benefits wealth estimates reported in this data set do not reflect WEP and/or GPO adjustments because we do not have sufficient information about pension benefits in non-covered jobs for all respondents.

After we calculate early, full, and delayed monthly retirement benefits for each of respondent in the matched sample, we adjust these numbers to 2016 U.S. dollars based on the intermediate economic assumption in the 2016 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 Social Security Benefit wealth in Section 4.

TABLE 1: Economic Assumptions

Wave Year	Nominal Interest	Inflation	Real Interest
1992	6.3%	4.0%	2.3%
1994	6.3%	4.0%	2.3%
1996	6.3%	4.0%	2.3%
1998	6.3%	3.5%	2.8%
2000	6.3%	3.3%	3.0%
2002	6.0%	3.0%	3.0%
2004	5.8%	2.8%	3.0%
2006	5.7%	2.8%	2.9%
2008	5.7%	2.8%	2.9%
2010	5.7%	2.8%	2.9%
2012	5.7%	2.8%	2.9%
2014	5.6%	2.7%	2.9%
2016	5.3%	2.6%	2.7%

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

## 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 their earnings history based on equation (1) 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 is still 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: `RwIMPUTE` and `SwIMPUTE`. We define these indicator variables and the corresponding imputation strategies in the following section. While the imputed wealth amounts are included in the public release of this dataset, the variables `RwIMPUTE` and `SwIMPUTE` are only available as restricted data.

### 3.2.1 Respondent Imputations

The indicator variable for respondent imputations, `RwIMPUTE`, 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 Benefit (self-reported);
- 3, if the respondent is unmatched and does not expect to claim Social Security Benefit or does not report expectations according to the self-report in the HRS survey;
- Missing, if the respondent is already receiving retirement benefits according to MBR (for the matched sample) or the HRS survey (for the unmatched sample).

In cases where `RwIMPUTE` 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 the respondent's own PIA at the FRA (calculated in the previous section) on several

demographic and labor force characteristics as of 2016.<sup>34</sup> To allow the coefficients in the imputation equation to differ by sex 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) 2010; (b) females first interviewed after 2010; (c) males first interviewed in or before 2010; and (d) females first interviewed in or before 2010. 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 2014, 2012, and 2010.

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.<sup>5</sup> 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 the unmatched respondent. If an unmatched respondent has more than one such “nearest neighbor” (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 accordingly. To calculate benefits wealth, we use the imputed PIAs as if they are the

---

<sup>34</sup>We include the following explanatory variables: respondent’s age, three marital status indicators (married, divorced, or widowed), eight regions of residence dummy variables, two dummy variables for race/ethnicity (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 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 wages for the current or last job, logarithm of total household income, occupation and industry indicators for the current or last job, and whether and how many years a respondent worked in jobs not covered by Social Security. 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.

<sup>4</sup>In the previous release of this data set, we included “the number of years in federal jobs before 1984” in the imputation equation because federal employees were not covered by Social Security before 1984; this variable is not in the 2016 HRS and hence excluded from the imputation equation in this release. Instead, we use whether and how many years a respondent worked in jobs not covered under Social Security (L110 and L111 in Section J3).

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



actual PIAs (see Section 4 for more details).

### 3.2.2 Spouse Imputations

In order to calculate benefits at the household level, we also need to impute the spouse's PIA if the spouse is not in the matched sample. The indicator variable for spouse imputations, SwIMPUTE, is defined very similarly to RwIMPUTE. 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 RwIMPUTE is the second member's SwIMPUTE, and the the first member's SwIMPUTE is the second member's RwIMPUTE. The only difference between RwIMPUTE and SwIMPUTE is that SwIMPUTE 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 Benefit in this dataset. A retired worker age 62 or higher who is fully insured is eligible to receive monthly retirement benefits (also called old-age 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 their spouse's earnings history (see §305 of the SSA Handbook). In this case, a fully insured retired worker receives a monthly benefit that is comprised of both 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 Cam and Sam are married. Cam gets \$200 per month for the SSA old-age benefits based on own work history. Sam gets \$1,000 per month for the old-age benefits. Cam is entitled to an auxiliary spouse benefit in the amount of \$500 from Sam's account. Since the \$500 is greater than Cam's own retirement insurance benefit amount (\$200), Cam is dually entitled. Cam will receive \$200 against their account and an additional \$300 against Sam's. We refer to the \$200 portion of

this benefit as the “retirement insurance benefit” and the \$300 portion as the “incremental auxiliary spouse benefit.” Note that Cam actually receives this all in one check in the amount of \$500. If Cam’s own benefit were \$600 instead of \$200, Cam would not be termed dually entitled and would have no auxiliary spouse benefit. Similarly, Sam is not dually entitled on Cam’s account because Sam is only eligible for half of Cam’s benefit (\$100), which is less than what Sam receives on their account. Table ?? which follows, illustrates how we define the types of benefits for this case for the construction of wealth estimates in this data set.

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
Cam	\$200	\$300
Sam	\$1,000	\$0

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

TABLE 3

	Own Benefit	Spousal Benefit	Survivor Benefit
Cam	\$200	\$0	\$800
Sam	\$0	\$0	\$0

In our calculation, we separate a respondent’s benefits into three components: (1) the retirement benefit based on the respondent’s own earnings history, (2) the spousal benefit when both spouses 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 tables<sup>6</sup> and the relevant interest and inflation rates as projected from the *Trustee Report*, we calculate the present value of the stream of Social Security old-age benefits at each claim age. The expected discounted present value of benefits is given by:

$$PV_{T_0} = \sum_{t=T_0}^{119} B_t \cdot {}_tP_{2016} \cdot (1+r)^{-(t-T_0)} \quad (2)$$

where  $B_t$  is the annual benefit payment at age  $t$ ,  ${}_tP_{2016}$  is the probability of being alive at age  $t$  conditional on having survived through 2016, and  $r$  is the nominal interest rate.

We sum the 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 adjust  $PV_{T_0}$  from claim year dollars to 2016 U.S. dollars, which is what is reported in the data set. Note that we also adjust the first year of benefits to account for the fractional consumption of benefits in that year. For example, if a respondent starts receiving benefits in July, then we use one-half of the relevant  $B_{T_0}$  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 2016 will have missing values for `RwSSWRER` (wealth based on own earnings record and early retirement claim age) and `RwSSWRNR` (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 spouses claim at the same age to simplify the calculation.<sup>7</sup>

---

<sup>6</sup>We use expected survival probabilities up to age 119 based on the 2016 version of sex-specific SSA cohort mortality tables.

<sup>7</sup>We also assume that, if a couple is married in 2016, they stay married though the remainder of their lives. Furthermore, to be eligible for spousal benefits, the marriage needs to last at least 10 years. We use the length of

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

$$\begin{aligned}
SSWealth_i = & \sum_{t=T_0}^{119} B_{it} \cdot {}_tP_{2016,i} (1+r)^{-(t-T_0)} + \\
& \sum_{t=T_0}^{119} \max(0, 0.5 \cdot (B_{jt} - B_{it})) \cdot {}_tP_{2016,i} \cdot {}_tP_{2016,j} (1+r)^{-(t-T_0)} + \\
& \sum_{t=T_0}^{119} \max(0, B_{jt} - B_{it}) \cdot {}_tP_{2016,i} \cdot (1 - {}_tP_{2016,j}) (1+r)^{-(t-T_0)}
\end{aligned} \tag{3}$$

where  $B_{jt}$  is the annual retirement benefit amount for spouse  $j$  at age  $t$ ,  ${}_tP_{2016,j}$  is the probability spouse  $j$  is alive at age  $t$  conditional on having survived through 2016, and  $r$  is the nominal interest rate.

The first term represents Social Security Benefit 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 by the joint probabilities that only one survives. Note that incremental auxiliary spouse and survivor wealth are zero for the higher earner (by definition). The wealth estimates are reported in 2016 dollars.

#### 4.4 Total Household Wealth

To make these values comparable across individuals, household wealth is reported in 2016 U.S. dollars. 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), wealth based on the respondent's own earnings history (the first term on the right hand side) and wealth based on spousal and survival benefits (the sum of the second and third

---

marriage at the relevant claim age, rather than as of 2016, to determine such eligibility.

terms on the right hand side) are both rounded to the nearest \$100 and then top-coded at the top 2 percent of values.<sup>8</sup>  $SSWealth_i$  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 old-age 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). We flag these individuals using `RwCLAIMED` and `SwCLAIMED`. They will have missing values for all SS wealth prediction variables.

## 5 Distribution Information

The Prospective Social Security Wealth Measures of Pre-Retirees data are distributed in one data file at the respondent level with Social Security Benefit wealth estimates in 2016. The records in the data files are identified by HHID and PN. 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

---

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

- \*.txt for codebook files.

For example,

- SSWEALTH.da contains ASCII data.
- SSWEALTH.sas contains corresponding SAS program statements,
- SSWEALTH.sps contains corresponding SPSS program statements,
- SSWEALTH.do contains corresponding Stata do statements,
- SSWEALTH.dct contains corresponding Stata dictionary statements, and
- SSWEALTH.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 (SSWEALTH.txt) and the data description (this document).

## 5.1 Variable Naming Conventions

The variable names are slightly different from the patterns used in the RAND-HRS data. The first character indicates whether the variable refers to the respondent's own benefit determined by their own earnings history, "R"; the respondent's spousal/survival benefit, "S"<sup>9</sup>; or the household, "H". The second and third characters correspond to the wave to which the variable pertains: "10" for 2010 and "13" for 2016. All Social Security wealth estimates 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 2016, otherwise, we assign missing values if the respondent is older than a particular claim age. We also

---

<sup>9</sup>This is different from the RAND-HRS data, where the S- variables indicate the corresponding values for the spouse.

assign missing values if the respondent has already started claiming benefits as evidenced by being in MBR. Table 4 shows which wealth values we calculate for individuals of a given age in 2016.

TABLE 4

Age in 2016	Assuming Claim Age of:		
	Early (Age 62)	FRA	Late (Age 70)
70 and Above	–	–	–
NRA-69	–	–	X
63-NRA	–	X	X
62 or Younger	X	X	X

Table 5 lists the variables included in the data file and some brief descriptions.  $w$  is either 10 (the 2010 wave) or 13 (the 2016 wave).

TABLE 5: List of Variables

Variable Name	Brief Description
HHID	Household Identifier
PN	Person Number
$RwSSWRER$	Respondent’s benefit wealth based on own earnings history, assuming claim age at 62; calculated as in equation (2).
$RwSSWRNR$	Respondent’s benefit wealth based on own earnings history, assuming claim age at FRA; calculated as in equation (2).
$RwSSWRXA$	Respondent’s benefit wealth based on own earnings history, assuming claim age at 70; calculated as in equation (2).
$RwCLAIMED$	An indicator that equals 1 if the respondent is already receiving benefits.
$SwSSWRER$	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).
$SwSSWRNR$	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).
$SwSSWRXA$	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).
$SwCLAIMED$	An indicator that equals 1 if the respondent’s spouse is already receiving benefits.
$HwSSWRER$	Total household benefit wealth, assuming claim age at 62; calculated as in equation (4).
$HwSSWRNR$	Total household benefit wealth, assuming claim age at FRA; calculated as in equation (4).
$HwSSWRXA$	Total household benefit wealth, assuming claim age at 70; calculated as in equation (4).

## **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 `*.sps` extensions, those with SAS statements with `*.sas` extensions, 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:\sswealth\sas" and the data file is located in "c:\sswealth\my data", you can run the file as is. A SAS system file (`*.sas7bdat`) will be saved to directory "c:\sswealth\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:\sswealth\spss” and the data file is located in “c:\sswealth\DATA”, you can run the file as is. An SPSS system file (\*.sav) will be saved to directory “c:\sswealth\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:\sswealth\Stata” and the data file is located in “c:\sswealth\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 and 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](mailto: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: <https://hrs.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: [hrsquestions@umich.edu](mailto:hrsquestions@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.