Health and Retirement Study
Prospective Social Security Wealth Measures of Pre-Retirees
1992–2020

Data Description and Usage

Chichun Fang
Institute for Social Research
University of Michigan

Version 1
March 2024

Conditions of Use

By registering for access to Health and Retirement (HRS) Public Release data, the User agrees to all of the following:

- Make no attempts to identify study participants.
- Not to transfer HRS Public Release data to any third party other than staff or students for whom you are directly responsible except as indicated below.
- Not to allow others to use your username and password to access this site.
- To certify the destruction of any downloaded Public Release data file as well as any data files derived from the downloaded file when requested to do so by the HRS.
- To include the following citation in any research reports, papers, or publications based on Public Release data:

  In text:

  "The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan."

  In references:

  "Health and Retirement Study, ([insert Product Name]) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (year)."

- To include the following citation in any research reports, papers, or publications based on any Public Release data file tagged as "Early" or "Preliminary": "This analysis uses Early Release data from the Health and Retirement Study, ([insert Product Name]), sponsored by the National Institute on Aging (grant number NIA U01AG009740) and conducted by the University of Michigan. These data have not been cleaned and may contain errors that will be corrected in the Final Public Release version of the dataset."
- Please note that although it is necessary to include the above citations to the HRS in any publications, it is not necessary to associate your publications with the HRS grant in PubMed.
- Provide information regarding any research product (dissertation, thesis, journal article, book, book chapter, report, etc.) based on data obtained from the HRS by sending an electronic copy to hrspublications@umich.edu.
- Report immediately to the HRS at hrsquestions@umich.edu any disclosure of study participant identity as well as any discovery of flaws or errors in the data or documentation files.

Notify the HRS through use of the update function provided at this site or by electronic mail directed to hrsquestions@umich.edu of changes in your electronic mail address, postal address, telephone number, organizational affiliation or organizational status.

Contents

1 Introduction

In this documentation, we outline assumptions, calculations, and other details pertaining to the construction of Social Security benefits 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, 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 data product is public-use and consists of respondent-level Social Security benefits wealth estimates constructed using the HRS survey data as well as SSA's administrative records. Each respondent is uniquely identified by the concatenation of the Household ID and the Person Number, HHID and PN. This release covers 1992 through 2020 waves and replaces all other public-use Social Security benefits wealth data products released by the HRS.

In particular, we utilize information from HRS respondents who consented[1] to authorize the University of Michigan's Institute for Social Research to obtain administrative records from the SSA to calculate and impute Social Security benefits wealth. 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."

Two sets of SSA administrative records are utilized in this data product: Summary Earnings Records (SER) and Master Beneficiary Records (MBR). SER contain the amounts of yearly earnings covered by Social Security between as early as 1951 and the most recent year that the consent remained valid.[2] MBR contain information on the type(s) of benefits (old-age, spousal, survivor, and/or disability insurance) and monthly benefits amounts for those who have applied for or claimed these benefits.

According to the agreement between HRS and SSA, the wealth estimates provided in this

[1] The respondent is required to provide the full name, Social Security number (SSN), and date of birth (DOB) on the consent form, which has to be signed and dated before being returned to the HRS. SSA does not accept partial names, SSN, or DOB; forms with incomplete information or without a signature are not considered as valid consents.

[2] Prior to 2006, consents were only valid for the particular wave when they were given. Since 2006, consents have become prospective and remain valid for 6 (12) years for earnings (benefits) records. Due to the different types/timing of consent given, the end year of earnings records differs across respondents.

data product are prospective, which means we only calculate wealth estimates for respondent-wave observations that have not yet claimed old-age, spousal, or survivors benefits according to SSA's administrative records[3] or survey self-reports (when administrator records are not available).

Section 2 covers the procedures and assumptions used to calculate or impute monthly Social Security benefits. Subsections 2.1 and 2.2 discuss the matched and the unmatched samples, respectively. Section 3 describes how we use the various components of Social Security benefits to calculate Social Security benefits wealth.

2 Benefits Calculation

This data products include Social Security benefits wealth estimates for all HRS waves since 1992. In the following, we discuss how wave year t (t=1992, 1994, 1996, etc.) estimates are calculated without loss of generality. We repeat the same algorithm to calculate estimates for all waves covered in this data product. Other notations are explained as they are introduced.

2.1 Earnings Projection and Benefits Calculation for the Matched Sample

We use batch version of SSA's AnyPIA program (Office of the Actuary, v2023.2) 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 algorithms and our assumptions in this section.

The AnyPIA program determines eligibility and coverage based on each individual's earnings history. SSA's Summary Earnings Records provide earnings from jobs (including self-employment) covered by Social Security. For the rest of this documentation, the term "earnings" refers to "earnings covered by Social Security" unless otherwise specified. Earnings are available since 1951 and are top-coded at each year's FICA taxable maximum.[4]

[3] Each insured individual only has entries in the MBR upon applying for Disability Insurance (DI) or claiming old-age, spousal, and survivors (OASI) benefits. If a respondent has matched SSA earnings records but no OASI entries in the MBR, we assume s/he has not claimed these benefits.

[4] Although earnings between 1937 and 1950, if any, do factor into the calculation of Social Security benefits, pre-1950 earnings are largely irrelevant for the sample/years that this data product covers.

Assuming that SSA earnings records are available through year $t-1$, we project earnings beyond year $t-1$ following Mitchell, Olson and Steinmeier (2000) for individuals who have not yet attained the corresponding claim ages (62, full retirement age [FRA], or 70) in year t.[5] Earnings in the five-year period preceding the wave year t are indexed to $t-1$ levels using SSA's national Average Wage Index (AWI). The indexed wages are then averaged, with years $t-1$ through $t-5$ earnings given weights 5, 4, 3, 2, and 1, respectively.

The projected earnings for the wave year t can be expressed as:

$$Y_t = \frac{1}{15}\Big(1+CPI_t\Big)\Big(5\cdot Y_{t-1} + 4\cdot Y_{t-2}\cdot\frac{AWI_{t-1}}{AWI_{t-2}} + 3\cdot Y_{t-3}\cdot\frac{AWI_{t-1}}{AWI_{t-3}} + 2\cdot Y_{t-4}\cdot\frac{AWI_{t-1}}{AWI_{t-4}} + 1\cdot Y_{t-5}\cdot\frac{AWI_{t-1}}{AWI_{t-5}}\Big)$$

where Y_t represents the earnings in year t and $\frac{AWI_{t-1}}{AWI_{t-i}}$ is that inflates earnings in year $t-i$ to year $t-1$ dollars.

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.

If SSA earnings records are only available through a different year t' that is earlier than $t-1$[6], we first use the same projection algorithm to project annual earnings between years $t'+1$ and $t-1$, and then years t and beyond.

In cases where respondents have \$0 earnings in year $t-1$, we examine their self-reported employment status in year t. 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 in years t and beyond. In other words, if a respondent is

[5] For each wave year t in this data product, we use SSA earnings records through year $t-1$ and then project earnings in years t and beyond to calculate benefits for year t. Actual earnings records beyond year t, even if available, are not used. For example, to calculate benefits for the 1992 wave, we use SSA earnings records through 1991 and project the earnings in years 1992 and beyond. We choose this strategy to ensure the methodological consistency across waves, despite that HRS has SSA earnings records for some respondents through 2020 at the time this data product was produced.

[6] This happens when a respondent's consent expired, and the HRS has not yet obtained a reconsent. In such cases the HRS still retains SSA earnings records obtained when the consent was valid, but the records are no longer updated once the consent expires. For example, if a respondent consented in 1992 but never re-consent afterwards, HRS still retains this respondent's earnings through 1991.

working in year t, we assume the respondent does not quit the labor market before the relevant claim age. If a respondent is not working in year t, we assume the respondent does not re-enter the labor market.

Given the 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.

The 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 the "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 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).[7] 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.

2.2 Benefits Imputation for the Unmatched Sample

We impute PIA and benefits only for the respondents whom we do not have matched SSA earnings records at all. As previously stated, in cases where we have some (but not all) earnings records for

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

a respondent, we extrapolate the earnings history based on equation (1) and rely on AnyPIA to calculate PIA and benefits for this respondent.

2.2.1 Respondent Imputations

For each respondent in the unmatched sample, we impute the respondent's own PIA using the nearest neighbor matching method. Starting with the sample of matched respondents, we regress logarithm of the respondent's own PIA at the FRA (calculated in the previous step) on several demographic and labor force characteristics as of year t.[8][9] We estimate the same model separately by gender and by wave.

We then use the coefficients from the corresponding imputation equation to "predict" PIA for the entire (matched and unmatched) sample and sort all observations by their predicted PIAs. For each member of the unmatched sample, we find the matched respondent who has the closest predicted PIA and assign the actual PIA (and hence benefits at various ages [62, FRA, and 70]) of this matched respondent as the imputed PIA 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. To calculate benefits wealth, we use the imputed PIAs as if they are the actual PIAs (see Section 3 for more details), based on which the amount of benefits at various claim ages can be determined.

[8] 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, whether and how many years a respondent worked in jobs not covered by Social Security, and whether the respondent expects to received Social Security benefits. 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.

[9] 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 available in all waves of the 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 in the waves that these variables are available. (The variable names have changed multiple times over time. In 2016 and later, they are L110 and L111 in Section J3).

2.2.2 Longitudinal Consistency

If a member of the unmatched sample is matched to different members of the matched sample (and hence potentially very different wealth estimates) in different waves, it would seem as if that unmatched respondent experiences fluctuations in Social Security benefits as s/he ages. To avoid this problem, we "fix" the imputation in this data product using the following two-step process.

First, in each wave, we impute the PIAs using the algorithm described in the previous subsection for each respondent in the unmatched sample. Second, we only retain the imputation in the age 60/61 wave, or the wave most adjacent to the age 60/61 wave if there is no age 60/61 wave. We then use the imputed PIA in this wave to calculate benefits in all other waves for the unmatched respondent. Thai is, while each respondent in the unmatched sample has an imputed PIA in step 1 (we do so in order to enlarge the pools of both the matched and unmatched sample), we only use the imputed PIA from the wave most adjacent to the age 60/61 wave in step 2, with which we use to replace the imputed PIAs in all other pre-retirement waves for this respondent.

2.2.3 Spouse Imputations

Since each "spouse" is also a HRS respondent, there is no need to separately impute PIAs for "spouses" in the unmatched sample. In other words, each respondent in our sample has his or her own PIA (imputed or not), which we use to calculate old-age, spousal, and survivors benefits accordingly.

3 Wealth Calculation

3.1 Types of Benefits

We focus on three distinct types of Social Security benefits 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 survivors 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 for more than 10 years. Cam gets $200 per month for the SSA old-age benefits based on their 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 own 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 1 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 1

	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 their own work history 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." As Table 2 summarizes:

In our calculation, we separate a respondent's benefits into three components: (1) the retirement

TABLE 2

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

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 respondents who were never married, or who were never entitled to spousal/survivor benefits due to the duration of marriage.[10] We calculate the present value of these three separate streams of income adjusted by the appropriate survival probabilities.

3.2 Wealth Based on Benefits from Respondent's Own Earnings History

Using SSA's 2022 Cohort Life Table and the relevant interest and inflation rates as projected from the *Annual Report of the Board of Trustees of the OASDI Trust Funds* ("*Trustees 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:

$$PV_{T_0} = \sum_{a=T_0}^{119} B_a \cdot {}_aP_t \cdot (1+r)^{-(a-T_0)} \tag{1}$$

where B_a is the nominal annual benefit payment at age a, ${}_aP_t$ is the probability of being alive at age a conditional on having survived through wave year t, and r is the nominal interest rate.[11] Inflation and interest rates are adopted from the intermediate economic assumption in year t version of the *Trustees Report*, summarized in Table 3.

We sum the discounted benefits from the claim year ($a = T_0$) through the age of 119. We adjust this value by the probability of survival through age a and discount it back to claim age (T_0) U.S. dollars. We also adjust the first year of benefits to account for the fractional consumption of

[10] In order to be eligible for spousal benefits, the marriage has to last for more than ten years. One might also be entitled for spousal benefits from a previous marriage that lasted for more than ten years. In this data product, the benefits wealth as of wave year t does not account for the benefits entitled from previous marriage that ended prior to year t interview, regardless whether this previous spouse as of year t is a HRS respondent or not.

[11] Since the real terms of Social Security benefits remain constant after claiming, this calculation can also be done in real terms as of age T_0, by setting the numerical values of B_a as B_{T_0} for all ages (between T_0 and 119) and discounting with real interest rates.

TABLE 3: 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%
2018	5.3%	2.6%	2.7%
2020	4.7%	2.4%	2.3%

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 use yearly compounding since our mortality data are at yearly intervals. To make these values comparable across individuals, we then further discount PV_{T_0} from claim year dollars to wave year t U.S. dollars using the corresponding nominal interest rate as of year t in Table 3.

3.3 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.[12]

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

[12]We also assume that a couple stays 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 marriage at the relevant claim age, rather than as of the wave year t, to determine such eligibility.

(assuming adjustments for early/late claiming have already been made to Bs):

$$
\begin{aligned}
SSWealth_i \quad = \quad & \sum_{a=T_0}^{119} B_{it} \cdot {}_aP_{t,i}(1+r)^{-(a-T_0)} + \\
& \sum_{a=T_0}^{119} max(0, 0.5 \cdot (B_{ja} - B_{ia})) \cdot {}_aP_{t,i} \cdot {}_aP_{t,j}(1+r)^{-(a-T_0)} + \\
& \sum_{a=T_0}^{119} max(0, B_{ja} - B_{ia}) \cdot {}_aP_{t,i} \cdot (1 - {}_aP_{t,j})(1+r)^{-(a-T_0)}
\end{aligned} \tag{2}
$$

where B_{ja} is the nominal annual retirement benefit amount for spouse j at age a, ${}_aP_{t,j}$ is the probability spouse j is alive at age a conditional on having survived through wave year t, and r is again 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 (1). The second term represents wealth based on the incremental auxiliary spousal benefit, which is adjusted by the likelihood that both spouses are alive in 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 benefits wealth are zero for the higher earner (by definition). As previously stated, the wealth estimates are further discounted from claim age T_0 to wave year t using the nominal interest rate as of year t, and are reported in wave year t nominal dollars.

3.4 Total Household Wealth

For members i and j in the same household, total household wealth[13] is defined as:

$$HHSSWealth_{ij} = SSWealth_i + SSWealth_j \tag{3}$$

3.5 Rounding and Top-Coding

In equation (2), wealth based on the respondent's own earnings history (the first term on the right hand side) and wealth based on spousal and survivor 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

[13] In a household where one member has claimed while the other has not, the household wealth only includes the wealth from the member who has not claimed.

2 percent of values.[14] $SSWealth_i$ is then calculated based on these rounded and top-coded terms. Neither $SSWealth_i$ in equation (2) nor $HHSSWealth_{ij}$ in equation (3) is further rounded and/or top-coded.

3.6 Wealth for Respondents Who Have Claimed

Individuals are designated as already having claimed if they have started receiving old-age, spousal, or survivors 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). They will have missing values for all Social Security benefits wealth estimates variables.

4 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 benefits wealth estimates in waves 1992 through 2020. 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:

- `SSW19922020.da` for data files,
- `SSW19922020.sas` for SAS program statements,
- `SSW19922020.sps` for SPSS program statements,
- `SSW19922020.do` for Stata do statements,
- `SSW19922020.dct` for Stata dictionary statements, and
- `SSW19922020.txt` for codebook files.

[14] 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.

The Social Security benefits 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 (`SSW19922020.txt`) and the data description (this document).

4.1 Variable Naming Conventions

The variable names follow the conventions used in the RAND-HRS data. The first character indicates whether the variable refers to the respondent's own benefits, "R"; the spouse's benefits, "S"[15]; or the household benefits, "H"[16]. The second and third characters correspond to the wave to which the variable pertains: '1' for the 1992 wave, '2' for the 1994 wave, etc. All Social Security wealth estimates are then followed by "SSW". The seventh character denotes the source of the benefit: "R" for benefits based on own earnings history, and "S" for spousal/survivor benefits[17]. 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.

In this data product, the assignment of "spouse" is based on the wave-specific *PPN variable (spouse/partner person number) in the HRS Tracker file. A "household" is determined based on the combination of HHID (household ID) and wave-specific *SUBHH (sub-household ID). These variables are directly available in the HRS Tracker file.[18] We do not correct for the longitudinal consistency of marital/partnership status. In the very few cases where *PPN and *SUBHH provide conflict information whether two respondents of the same HHID are a couple and hence in the same household, the summation of corresponding "R" and "S" variables (the sum of benefits between the respondent and the spouse) may not equal to the "H" variable (total household benefits).

We calculate benefits assuming claim ages that the respondent has not yet attained as of wave year t, otherwise, we assign missing values if the respondent is older than a particular claim age.

[15]The "S" variables are set to missing for unmarried respondents.

[16]As previously mentioned, the "H" variables only include the wealth of the household member(s) who had not claimed OSAI benefits as of the corresponding wave.

[17]Spousal/survivor benefits are set to missing for unmarried respondents, for respondents who do not have long enough marriage to be eligible for such benefits at the corresponding claim age, and for respondents whose spouses do not provide sufficient information for the HRS to calculate or impute their PIAs (mostly due to incomplete interviews).

[18]The * indicates the variables are wave specific, and there is a wave indicator preceding the variable name. For example, the spouse/partner person number in the 1992 wave is APPN.

We also 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 wave year t. "Y" indicates the wealth estimates are available while "–" indicates missing values.

TABLE 4

	Assuming Claim Age of:		
Age in wave year t	Early (Age 62)	FRA	Late (Age 70)
70 and Above	–	–	–
FRA-69	–	–	Y
63-FRA	–	Y	Y
62 or Younger	Y	Y	Y

Table 5 lists the variables included in the data file and some brief descriptions. w ranges from 1 (1992 wave) to 15 (2020 wave). All the wealth variables are in wave w nominal US dollars. The rules in Table 4 apply to all the wealth variables.

4.2 Identification Variables

4.2.1 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.

4.2.2 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.

4.3 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.

4.3.1 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:\SSW19922020\sas`" and the data file is located in "`c:\SSW1992202\data`", you can run the file as is. A SAS system file (`*.sas7bdat`) will be saved to directory "`c:\SSW19922020\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.

4.3.2 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:\SSW19922020\spss`" and the data file is located in "`c:\SSW19922020\DATA`", you can run the file as is. An SPSS system file (`*.sav`) will be saved to directory "`c:\SSW19922020\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.

4.3.3 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:\SSW19922020\Stata`" and the data file is located in "`c:\SSW19922020\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.

4.4 Registration and Downloading the Data

HRS data are available for 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 e-mail. Your username and password are required to download data files.

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.

By registering, you agree to the Conditions of Use (https://hrsdata.isr.umich.edu/data-products/conditions-of-use) governing access to the HRS public release data.

4.4.1 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.

4.4.2 Publications Based on the HRS 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 an electronic copy of any publications you produce based on HRS data, with a bibliographical reference, if appropriate, to hrspublications@umich.edu.

4.5 If You Need to Know More

This document is intended to serve as a brief overview and to provide guidelines to using the pension wealth estimates 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.

4.5.1 HRS Website

HRS 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

4.5.2 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
- Postal Service:

  Health and Retirement Study

  Institute for Social Research

  University of Michigan

  P.O. Box 1248

  Ann Arbor, MI 48106-1248

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 Euller, Cynthia Grant, Melissa Bradley, Christin E. Peterson, Randall Hirscher, and Paul Steinberg. 2000. *SSA Program Data User's Manual.* RAND.

TABLE 5: List of Variables

Variable Name	Brief Description
HHID	Household Identifier
PN	Person Number
R*w*SSWRER	Respondent's benefits wealth based on own earnings history, assuming claim at age 62; calculated as in equation (1).
R*w*SSWSER	Respondent's benefits wealth based on spousal and survivors benefits, assuming claim at age 62; sum of the second and third components in equation (2).
R*w*SSWRNR	Respondent's benefits wealth based on own earnings history, assuming claim at FRA; calculated as in equation (1).
R*w*SSWSNR	Respondent's benefits wealth based on spousal and survivors benefits, assuming claim at FRA; sum of the second and third components in equation (2).
R*w*SSWRXA	Respondent's benefits wealth based on own earnings history, assuming claim at age 70; calculated as in equation (1).
R*w*SSWSXA	Respondent's benefits wealth based on spousal and survivors benefits, assuming claim at age 70; sum of the second and third components in equation (2).
R*w*CLAIMED	An indicator that equals 1 if the respondent has already claimed OASI benefits.
S*w*SSWRER	Spouse's benefits wealth based on own earnings history, assuming claim at age 62; calculated as in equation (1).
S*w*SSWSER	Spouse's benefits wealth based on spousal and survivors benefits, assuming claim at age 62; sum of the second and third components in equation (2).
S*w*SSWRNR	Spouse's benefits wealth based on own earnings history, assuming claim at FRA; calculated as in equation (1).
S*w*SSWSNR	Spouse's benefits wealth based on spousal and survivors benefits, assuming claim at FRA; sum of the second and third components in equation (2).
S*w*SSWRXA	Spouse's benefits wealth based on own earnings history, assuming claim at age 70; calculated as in equation (1).
S*w*SSWSXA	Spouse's benefits wealth based on spousal and survivors benefits, assuming claim at age 70; sum of the second and third components in equation (2).
S*w*CLAIMED	An indicator that equals 1 if the spouse has already claimed OASI benefits.
H*w*SSWRER	Total household benefits wealth, assuming claim age at 62; calculated as in equation (3).
H*w*SSWRNR	Total household benefits wealth, assuming claim age at FRA; calculated as in equation (3).
H*w*SSWRXA	Total household benefits wealth, assuming claim age at 70; calculated as in equation (3).