INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
HEALTH AND RETIREMENT STUDY
UNIVERSITY OF MICHIGAN

# Health and Retirement Study
# Imputations for Employer-Sponsored Pension Wealth from Current Jobs in 2016

## Data Description and Technical Documentation

Chichun Fang
Institute for Social Research
University of Michigan

Version 1.0
March 2022

*Pension plan information used to produce this data product is collected by a team led by Amy Butchart and Julia Roach, with student research assistants Tamer Abuhalawah, Colin Case, Kevin Chien, Yenibel Cuevas, Hannah Dang, Sylvie Evarts, Apelete Houngbo, Scott Kim, Courtney Laudick, DewRina Lee, Max Shipband, Becca Su, Yanling Xu, and Jalen Zeman. Helena Stolyarova also provided valuable inputs.

# 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 Overview

The *Imputations for Employer-Sponsored Pension Wealth from Current Jobs in 2016* data release consists of information derived from 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 HRS is a rich source of longitudinal data for researchers and policy makers who study aging. The National Institute on Aging provided funding, with supplemental support from the Social Security Administration. The study is conducted by the Institute for Social Research at the University of Michigan.

The HRS began with two separate studies: the original HRS cohort, born between 1931 and 1941, first interviewed in 1992 (at 51-61 years of age); and the AHEAD (Asset and Health Dynamics among the Oldest Old) cohort, born before 1923 and first interviewed in 1993 (ages 70 and above). Both samples included spouses or partners, regardless of age.[1] In 1998, after two follow-ups of HRS (1994, 1996), and one of AHEAD (1995), the studies were merged and combined with new respondents born 1942-1947 (War Baby, WB) and 1924-1930 (Children of the Depression Age, CODA) to become a complete panel of the population over age 50. New ample cohorts of individuals born in 1945-53 (Early Baby Boomer, EBB), 1954-59 (Mid Baby Boomer, MBB), and 1960-65 (Late Baby Boomer, LBB) were added in 2004, 2010, and 2016, respectively. The new cohorts were always aged 51-56 when they were added to the HRS. Future plans call for a new 6-year birth cohort to be enrolled every six years, beginning in 2022. All cohorts are drawn as nationally representative samples. However, those based on the 1992 screen contained over-samples of African Americans and Hispanic Americans, and the 2010 cohort also included an expansion of the minority sample. For more information about the HRS, please refer to Juster and Suzman (1995) and Sonnega et al. (2014).

[1] The HRS sample came from a screening of 69,336 households conducted in 1992 and generated using a multi-stage, clustered area probability frame. The AHEAD sample was generated for what began as a separate study. Individuals born between 1914 and 1923, and about half of those born in 1913 or before, were identified through the same household screening used to identify the original HRS sample. The other half of those born in 1913 or before were identified using the Medicare enrollment files maintained by the Health Care Financing Administration (HCFA, since renamed the Centers for Medicare and Medicaid Services, or CMS).

## 2 Background

Pension wealth estimates can be constructed either from calculating the present discounted value of a stream of future benefits, or from a reported current account balance. For this version of pension wealth estimates, we follow previous practice for 1992, 1998, 2004, and 2010 in using the respondent's self-reported pension plan type to determine which method of wealth estimation to use.

There were 4,166 respondents who reported participating in a pension at their current employer and were not self-employed in the 2016 survey. Among them, 1,567 reported to have at least one defined benefit (DB) plan, and 2,898 reported to have balance in at least one defined contribution (DC) plan. For the purpose of pension wealth estimation, a combination (Combo) plan is counted both as a DB plan and a DC plan; we report both the wealth in the DB component as well as the balance in the DC component of the plan.

If a respondent reported being covered by at least one DB plan in 2016, the DB pension wealth value for each such plan was calculated by the HRS *Pension Estimation Program* using a combination of self-reported data from the employment section of the 2016 survey and pension plan rules obtained from the plan's Summary Plan Description (SPD). We detail how the SPDs were obtained and used in Section 3.1. If a respondent reported being covered by least one DC plan on his current job in 2016, his/her DC wealth was taken from the self-reported account balance in the 2016 survey.

Imputations were necessary when (a) the information in the SPD was insufficient (or there was no SPD at all) for the respondent's self-reported DB pension plan, or (b) a respondent failed to provide an account balance for a self-reported DC plan in 2016. We discuss how DB wealth or DC account balances were imputed later in Sections 3.3 and 4.2, respectively.

# 3 Pension Wealth for Defined Benefit Plans

## 3.1 Obtaining Summary Plan Descriptions

We collected plan descriptions from various sources. For the remaining of this document, "project work" refers to the plan information that we collected from sources other than the HRS survey, and "self-report" refers to respondents' self-reports in the HRS survey.

For respondents in the private sector, we relied on Department of Labor's Form 5500 database[2] to obtain plan description documents. Form 5500 is an annual filing mandated by the Employee Retirement Income Security Act (ERISA), which is namely a plan-level report regarding its financial health. For large (more than 100 participants) plans, descriptions of eligibility, vesting, and benefit accrual rules are usually attached to the Form 5500 filing.

Public sector plans are not covered by ERISA; hence, they do not have corresponding Form 5500 filings. However, public plan SPDs are usually available on the Internet. We located the plan SPDs by directly going to the website of the plan sponsor.

Employers were identified using respondents' survey reports of employer names, addresses, and phone numbers. Each report was individually reviewed, and web searches and consultation of the business directory ReferenceUSA were used to determine accurate local and parent employers. Multiple criteria were assessed including employer name, location, industry, and phone. In addition, fuzzy match algorithm (Wasi and Flaaen, 2015) was used to compare self-reports in the survey to employers in the database of Department of Labor Form 5500 filings and in an internal database of past survey reports of employers. Employer matches were found for 4,040 of the 4,166 HRS respondents in 2016 who reported participating in a pension at their current employer and who were not self-employed.

Retirement plans in the Form 5500 database can be uniquely identified by the sponsor EIN and plan number. Once the employer was identified, we looked through all the plans sponsored by this employer for the plans that the HRS respondent may be eligible for. We call a plan document "codable" if it contains sufficient information to determine whether a respondent is eligible for this

[2] A web-based interface of this database is located at https://www.efast.dol.gov/5500Search/.

plan, when the benefit will vest, and how the benefit is calculated upon retirement or separation. After we found a codable plan document for a certain respondent, the plan was "coded" in the project work and the link between the respondent and the plan was established. Information in the documents that we found may not be consistent with the respondent's self-report, and sometimes we were not able to find enough information to establish the respondent-plan link. Such scenarios are discussed in section 3.3.

Given that our purpose is to calculate wealth in defined benefit plans, we only coded the information in the DB components of the plans that the HRS respondent may be eligible for.[3] If the matched plan had a DC component, it was recorded that the respondent had a DC plan. However, the provision in a DC plan was not coded.

## 3.2 Estimating Benefit in Defined Benefit Plans

*Pension Estimation Program* was used to calculate the present values of benefits from Defined Benefit (DB) plans for respondents linked to a pension plan in 2016. If that link was missing, a link was imputed to link the respondent to an actual plan (see Section 3.3). Years of service and earnings are two of the most important factors that determine the level of benefit upon retirement. Below we briefly explain how these measures are calibrated.

### 3.2.1 Hire Date and Years of Service

In calculating the benefits due from a defined benefit plan, participants' hire date is an important measure, as it relates to both plan eligibility and service credit. HRS respondents are surveyed repeatedly concerning the date of hire at their current employer, but these measures can be reported inconsistently. In this work, inconsistent reports of hire date were compared to Social Security earnings records when they were available. When earnings records could not be used to resolve inconsistent reports, the earliest reported hire date was used to determine plan eligibility, while service credit was calculated from the latest reported hire date.

[3] Account balance in defined contribution plans depends on both cumulative contributions (from employee and the employer) and historical returns. Due to incomplete knowledge about a respondent's past contributions and portfolios, we did not calculate DC account balance in the project work. For DC plans, we relied on the self-reported account balance in the HRS.

### 3.2.2 Earnings

Plan participants' earnings is another critical component of most defined benefit pension formulas. In the base case, RAND data were used to calculate earnings by multiplying together measures of weeks worked per year, hours worked per week, and hourly wage.[4] If the hourly wage was missing, a series of rules were used to select and apply substitute measures of income.

If the respondent was paid by salary, these substitutions were made from yearly income data reported in the employment section in the same interview wave. If the respondent was paid piece-work or "other", the substitution was made from the same wave income section data.[5] If this data were not available, income was calculated from adjusted employment section measures of income in adjacent interview waves for respondents with the same job in the relevant wave. Failing that, the current interview wave's income section data regarding the past year's income were used if the reported hire date was before 2015. If none of these measures were available, income was imputed to be the median income of respondents paid in that method (salary, hourly, or "other").

Whenever the information from a different wave/year was used to fill in as the earnings in 2016, we adjusted these earnings to 2016 U.S. dollars using the following rule:

- For the 2014 earnings, we applied a 4% increase for government sector and a 2.5% increase for private sector;
- For 2018 wage, we applied a 3% decrease for government sector and a 4% decrease for private sector.

The motivation behind these rules, rather than using the wage growth rate assumed in the *Annual Report of the Board of Trustees of the OASDI Trust Funds*, is that we want to account for the earnings growth at the individual level. Since people tend to leave near their peak earnings and are replaced in the labor force by people at lower earnings on their life cycle trajectories, the in-sample

[4] In about 40 cases where the hours worked were missed, we used the hours worked reported in 2014 survey if the respondent had the same employer and position in 2014. If the respondent had different employers or positions in all three waves, we assumed the hours worked in 2016 was 2,080.

[5] The wage income reported in the income section of the 2016 survey was the annual labor earnings in 2015. We only used wageincome from the income section when we were positive that the earnings was from the same job as the self-reported "current" job as of 2016.

wage growth (which is the earnings growth at the individual level) is in theory more appropriate than the SSA wage index or the long-term earnings growth rate assumed in the OASDI actuarial report (both are earnings growth of the whole population) for our purpose. We obtain the in-sample wage growth using the average earnings changes among the HRS respondents who stayed in the same jobs and had no missing earnings information between 2014 and 2018. The growth rates are calculated by sector to reflect different earnings trajectories between public and private sectors during and after the Great Recession.

### 3.3 Imputation of Benefit in DB Plans

When a respondent reported in the survey being covered by a DB plan but did not have a matched plan in the project work, this respondent became a "seeker"– the respondent needed plan benefits imputed. For each "seeker", we assigned a plan from the "donor" pool as this respondent's imputed DB plan. Table 1 shows the composition of donor and seeker pools based on both survey response and project work.

TABLE 1: Donor versus Seeker of DB Plans

| Self-Report (PJ338) | Project Work | | | |
|---|---|---|---|---|
| | DB Only | DC Only | Both or Combo | Not Matched |
| DB Only | Donor | Seeker | Donor$^{\dagger}$ | Seeker |
| DC Only | Donor$^{\ddagger}$ | N/A | Donor$^{\dagger\ddagger}$ | N/A |
| Both/Combo | Donor | Seeker | Donor$^{\dagger}$ | Seeker |
| DK/RF | Seeker | N/A | Seeker | Seeker |

$^{\dagger}$: Only the DB component of the plan is used as a donor plan.
$^{\ddagger}$: The coded plan from project work is used as a donor, but the respondent will not be assigned a DB wealth from this plan.

More specifically, we put all the respondents into bins constructed based on a hierarchy of sector (two categories: public versus private), union status (two categories: covered by a collected bargaining contract or not), education (five categories: less than high school, GED, high school, some college, and college or above), industry (19 categories), occupation (25 categories), size of the parent company (4 categories: less than 25, 25 to 99, 100 to 499, and 500 or more), and tenure in the current job (6 categories: less than 5 years, 5 to 10 years, 10 to 15 years, 15 to 20 years, 20 to

25 years, and 25 years and above).

For the bins that had at least 5 respondents and at least a donor, each seeker in these bins took a random draw (with replacement) of donors in the same bin. The drawn donor plan hence became the seeker's "imputed" plan. Bins that had less than 5 respondents or no donors were aggregated "upwards" in the hierarchy until they had at least 5 respondents and a donor, and seekers in these bins were assigned donor plans in the same manner.

Finally, we used a different set of rules for respondents (a) who had the same job in 2010 and 2016 and (b) whose plan linkage had to be imputed in both 2010 and 2016. In theory, such respondents should have the same plans in both waves. The imputation algorithm introduced previously does not account for cross-wave consistency, so re-imputing plan linkage in 2016 likely will give them different plans across waves. Instead, we utilized the results of the 2010 imputation for these respondents. If an imputed 2010 plan was also coded in 2016, we assigned the same plan as the 2016 imputed plan. If an imputed 2010 plan was not coded 2016, we calculated pension wealth by feeding the 2016 individual/job characteristics to the imputed 2010 plan. Hence, the changes in pension wealth for such respondents would reflect the changes in plan provision (for example, a plan might have been frozen during this period) or individual/job characteristics between 2010 and 2016, but not the change in imputed plan linkage across waves.

We used the variable "DBImputed" to indicate whether the person-plan link was imputed. This variable took three values:

- 0, if a matched plan was found for this respondent;
- 1, if an imputed person-plan link (and hence imputed wealth) was assigned to this respondent;
- Missing, if the respondent did not report to have a DB plan.

### 3.4 Present Value Calculation

The present value of benefit wealth from a DB plan was calculated as:

$$PV_{T_0} = \sum_{t=T_0}^{119} {}_tP_{2016} \cdot \Big(\frac{1+COLA}{1+r}\Big)^{(t-T_0)} \cdot B_{t|T_0} \tag{1}$$

where $PV_{T_0}$ is the present value of benefit wealth at retirement age (or quit date) $T_0$, ${}_tP_{2016}$ is the probability of being alive at age $t$ conditional on having survived through 2016[6], $COLA$ (cost-of-living adjustment) is the plan-specific annual growth rate of payment, $r$ is the nominal interest rate, and $B_{t|T_0}$ is the annual pension benefit at age $t$ conditional on retiring or quitting at $T_0$.

DB wealth estimates were calculated or imputed for each respondent at various retirement ages ($T_0$), which we detail in Section 3.5. The calculations assumed a real interest rate of 2.7% and an inflation rate of 2.6% according to the intermediate economic assumption in the 2016 version of the *Annual Report of the Board of Trustees of the OASDI Trust Funds.* To facilitate the comparison, the present values obtained in equation (1) were then discounted or inflated to 2016 U.S. Dollars.

### 3.5 Various Timing of Retirement

Present values of pension wealth were calculated or imputed at eight different points of time: 60, 62, 65, 70, early retirement age, normal retirement age, expected retirement age, and right now (that is, year 2016).[7]

At any given age, the *Pension Estimation Program* has three sets of wealth PVs: PV of early retirement benefits per plan language, PV of normal retirement benefits per plan language, and PV of benefits that are already vested (i.e. "vested and deferred").[8] In a given specific age-year, the PVs at early/normal retirement age would be zero for the years when the respondent was not eligible for early/normal benefits. Similarly, the PV of vested and deferred benefits would be zero if the benefits were not vested yet. In our calculation, we defined the maximum among these three numbers at a given age as the wealth at that age. The reason behind this is intuitive: when the retirement benefits were vested, we assumed that respondents would choose the provision that yielded the largest benefit payments.

The wealth estimate at age 60 is the present value of all future cash flows that the respondent

[6]We use expected survival probabilities up to age 119 based on the 2016 version of gender-specific SSA cohort mortality tables.

[7]In our calculation, we assume that each respondent quits on his/her birthday of the given age/year. To preserve confidentiality, the HRS only releases the birth year and month, so we further assume that HRS respondents are all born on the fifteenth day of their respective birth months. For example, the "right now" calculation provides pension wealth estimates assuming each respondent stops working on the fifteenth day of his/her birth month in 2016.

[8]Not all the plans have these three provisions. For example, a plan may not provide early retirement benefit.

will receive if she or he retires on the $60^{th}$ birthday. For the respondents younger than 60, such definition is straightforward. For the respondents older than 60, however, the wealth at age 60 is defined as the present value of all the future cash flows from 2016 onward assuming that the respondent has retired at age 60. In other words, we excluded all the benefits that had been paid in our calculation. Wealth at age 60 would be missing if the respondent was older than 60 when the current job started, and it would be zero if the benefits were not vested yet at age 60. We defined and calculated the wealth estimates at ages 62, 65, and 70 similarly.

PV at the early retirement age is the benefit wealth when a respondent initially becomes eligible for early retirement benefit. Empirically, in a *Pension Estimation Program* output that chronologically lists the benefits at each retirement/quit age, this would the first age when the PV at early retirement age becomes non-zero. We also provided the corresponding age in this dataset. Note that the early retirement age is defined at the person-plan level. It could differ across respondents in the same plan if, for example, the retirement age eligibility is determined by a combination of age and years of service. We defined and calculated the PV at normal retirement age in a similar manner.

PV at the expected retirement age is the pension wealth at the self-reported expected retirement age in the survey. If the self-reported expected retirement age was missing, we imputed it using the nearest neighbor matching. If the expected retirement age was smaller than the respondent's age in 2016, we assumed that the respondent expected to retire at the end of 2016. In other words, we assumed expected retirement age to be larger than or equal to the current age as of 2016. For respondents who were younger than 80 as of 2016, we also capped (top-coded) the expected retirement age at 80. For the respondents who were older than 80 as of 2016, we assumed they expected to retire in the end of 2016. Note that, if the benefits were not vested as of the expected retirement age, the PV at expected retirement age would be zero. We also provided the expected retirement age (modified, if necessary, according to the above rules) as well as an indicator of whether the expected retirement age was imputed in this dataset.

Finally, following Gustman, Steinmeier, and Tabatabai (2010*a*; 2010*b*), we define PV in 2016

(i.e. retiring right now) as a proration of wealth PV at the expected retirement age based on the ratio of the respondent's years of service as of 2016 to the respondent's expected years of service at the expected retirement age. Hence, it is the linearly-approximated portion of benefits at expected retirement date that the respondent has earned based on years of service as of 2016. Although most of the DB plans are by construction back-loaded, respondents tend to have "targeted" retirement dates and hence corresponding levels of benefits. As a result, for the purpose of evaluating retirement preparedness, a proration of benefits wealth at expected retirement age is a more appropriate measure of retirement wealth than the PV of wealth at the age in 2016.

# 4 Pension Wealth for Defined Contribution Plans

## 4.1 Account Balances in Defined Contribution Plans

To construct pension wealth from Defined Contribution (DC) plans, we used self-reported account balances from the current job as of the 2016 survey. A respondent could report multiple accounts balances from the same job. The total DC wealth was computed as a sum of all accounts from current job in 2016. If some of these amounts were missing, they were imputed (see Section 4.2).

## 4.2 Imputation of Account Balance in DC Plans

In this dataset, a respondent has a DC account balance when at least one of the self-reported plan types is DC or Combo. The DC account balance is pulled from Section J in 2016 of the HRS. When the account balance was not available, we imputed the respondent's DC account balance using a variation of the nearest neighbor matching method detailed below. The DC account balance is reported at the respondent level in this dataset.

Recall that the respondent could report an actual number or a "bracketed" range (for example, more than $20,000 and less than $50,000) in the DC account balance. We first aggregated the account balance up to the respondent level for those who had more than one DC/Combo account, accounting for both actual and bracketed answers. For example, if a respondent reported $100K and $25K in each account, her total balance was $125K. If a respondent reported a balance of $120K in the first account and "between $20K and $50K" in the second, she was considered to have a

bracketed answer between $140K and $170K. If the report was "between $20K and $50K" in the first account and "between $0 and $20K" in the other, the balance was between $20K and $70K. If the reports wer "between $20K and $50K" and "DK" in each account, the combined balance was "more than $20K."

Starting with the respondents whose DC account balances are actual numbers (that is, neither in bracketed numbers nor missing), we regressed the logarithm of respondent's DC account balance in 2016 on several socioeconomic, employment, and demographic characteristics as of 2016.[9] Based on the coefficients from this regression, we calculated a "predicted" account balance for all respondents, including those whose answers were actual numbers, bracket numbers, or missing. The sample was then sorted by this predicted account balance.

In the next step, a "nearest neighbor" was found for each of the respondent whose reported DC account balance was in brackets or missing. The actual account balance of the nearest neighbor was assigned as the imputed account balance for a respondent whose report was in brackets or missing. For a respondent whose reported DC account balance was missing, finding a nearest neighbor is straightforward. It was simply the respondent who reported an actual number in the account balance and had a predicted account balance adjacent to that of the respondent who needed an imputed balance. For a respondent whose report balance was in brackets, it was the most adjacent respondent whose actual account balance fell in the same reported bracket of the respondent who required imputation. That is, the "nearest neighbor" of a respondent who had a reported balance of "between $20K and $50K" was the respondent with the most adjacent predicted balance among those who had an actual balance between $20K and $50K.[10] In this way, the information in bracketed answers was preserved in the imputation process.

[9]We 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., years of work experience, a union status indicator, a dummy variable indicating whether there is a supplemental DB plan, a health insurance indicator, a home ownership indicator, logarithm of annual wage for the current job, logarithm of total household income, and occupation and industry indicators for the current job. For the logarithm of account balance, 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 account balance, earnings, or income are not dropped.

[10]In other words, a missing balance can be thought as a special case where the reported bracket is between $0 and the maximum amount reported in the survey.

We used the variable "ImpDCBal" to indicate whether the DC account balance was imputed. This variable took three values:

- 0, if the actual DC account balance was obtained from the self-report for this respondent;
- 1, if the respondent was assigned with an imputed DC balance;
- Missing, if the respondent did not report to have a DC plan.

# 5 Distribution Information

There are two data files, both at the respondent level, in the distribution package. The *Imputations for Employer-Sponsored Pension Wealth from Current Jobs in 2016* data are distributed in one data file (`PWI16A_R`). This file has all the information for the users who prefer to use wealth estimates released by the HRS rather than making their own imputations and calculations. The other file, `PWI16B_R`, contains the parameters that were fed into the *Pension Estimation Program*.

The records in the data files are sorted in order by the primary identifiers, 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,
- `*.do` for Stata do statements,
- `*.dct` for Stata dictionary statements, and
- `*.txt` for codebook files.

For example,

- `PWI16A_R.da` contains ASCII data.

- `PWI16A_R.sas` contains corresponding SAS program statements,
- `PWI16A_R.do` contains corresponding Stata do statements,
- `PWI16A_R.dct` contains corresponding Stata dictionary statements, and
- `PWI16A_R.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 (`PWI16A_R.txt`) and the data description (this document).

### 5.1 Variable Naming Conventions

Table 2 lists the variables included in the data file and some brief descriptions.

TABLE 2: List of Variables

| Variable Name | Brief Description |
|---|---|
| HHID | Household Identifier |
| PN | Person Number |
| HireDate | Year and Month of Hire |
| DBFlag_16 | Indicator that R Has DB Wealth |
| DBImputed | Indicator that DB Plan and Hence Wealth Is Imputed |
| PVage60_16 | DB Wealth at Age 60 |
| PVage62_16 | DB Wealth at Age 62 |
| PVage65_16 | DB Wealth at Age 65 |
| PVage70_16 | DB Wealth at Age 70 |
| PVageER_16 | DB Wealth at Early Retirement Age |
| AgeER_16 | Age Corresponding to PVageER_16 |
| PVageNR_16 | DB Wealth at Normal Retirement Age |
| AgeNR_16 | Age Corresponding to PVageNR_16 |
| PVExpRetAge_16 | DB Wealth at Expected Retirement Age |
| ExpRetAge_16 | Expected Retirement Age |
| ImpRetAge_16 | Indicator that Expected Retirement Age Is Imputed |
| PVage2016_16 | DB Wealth as of 2016 |
| Age2016_16 | Age as of 2016 |
| DCFlag_16 | Indicator that R Has DC Balance in 2016 |
| DCAcctBal | DC Account Balance as of 2016 |
| ImpDCBal | Indicator that DC Account Balance as of 2016 Is Imputed |

## 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`, 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:\pwi16\sas`" and the data file is located in "`c:\PWI16\my data`", you can run the file as is. A SAS system file (`*.sas7bdat`) will be saved to directory "`c:\pwi16\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 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:\pwi16\Stata`" and the data file is located in "`c:\pwi16\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 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.

### 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 publications you produce based on HRS data, with a bibliographical reference, if appropriate, to the address below.

Health and Retirement Study

Attn: Papers and Publications

The Institute for Social Research

University of Michigan

P.O. Box 1248

Ann Arbor, MI 48106-1248

Alternately, you may send an electronic copy to hrspublications@umich.edu.

## 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 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.

### 5.5.1 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

### 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
- Postal Service:

  Health and Retirement Study

  The 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.** 2010*a*. *Pensions in the Health and Retirement Study.* Harvard University Press.

**Gustman, Alan L., Thomas L. Steinmeier, and Nahid Tabatabai.** 2010*b*. "What the Stock Market Decline Means for the Financial Security and Retirement Choices of the Near-Retirement Population." *Journal of Economic Perspectives*, 24(1): 161–182.

**Juster, F. Thomas, and Richard Suzman.** 1995. "An Overview of the Health and Retirement Study." *Journal of Human Resources*, 30(S): S7–S56.

**Sonnega, Amanda, Jessica D Faul, Mary Beth Ofstedal, Kenneth M Langa, John WR Phillips, and David R Weir.** 2014. "Cohort Profile: the Health and Retirement Study (HRS)." *International Journal of Epidemiology*, 43(2): 576–585.

**Wasi, Nada, and Aaron Flaaen.** 2015. "Record Linkage Using STATA: Pre-Processing, Linking and Reviewing Utilities." *Stata Journal*, 15(3): 1–15.