

## **Health and Retirement Study**

**Prospective Social Security Wealth Measures of Pre-retirees  
Public Release  
Version 6.0, April 2022**

### **Data Description and Usage**

Prepared by Kandice Kapinos  
with Charlie Brown, Michael Nolte, Helena Stolyarova, David Weir

Survey Research Center  
Institute for Social Research  
University of Michigan  
Ann Arbor, MI

## Table of Contents

<b>I.</b>	<b>INTRODUCTION AND OVERVIEW .....</b>	<b>2</b>
<b>II.</b>	<b>BACKGROUND .....</b>	<b>2</b>
<b>III.</b>	<b>PROCEDURES.....</b>	<b>2</b>
A.	INITIAL CALCULATIONS AND ASSUMPTIONS.....	2
B.	WEALTH FOR RESPONDENTS WHO HAVE NOT YET CLAIMED .....	4
1.	<i>Background: Types of Benefits .....</i>	<i>4</i>
2.	<i>Wealth Based on Retirement Insurance Benefits.....</i>	<i>5</i>
3.	<i>Total Household Wealth.....</i>	<i>6</i>
C.	WEALTH FOR RESPONDENTS WHO HAVE CLAIMED .....	6
D.	VARIABLE NAMING CONVENTIONS.....	7
E.	VARIABLES FOR EACH WAVE.....	7
F.	DATA CODEBOOK .....	7
1.	<i>Rounding and Top-Coding .....</i>	<i>8</i>
G.	IMPUTATIONS .....	9
1.	<i>Respondent Imputations .....</i>	<i>9</i>
2.	<i>Spouse Imputations.....</i>	<i>10</i>
<b>IV.</b>	<b>DISTRIBUTION INFORMATION.....</b>	<b>11</b>
<b>V.</b>	<b>IDENTIFICATION VARIABLES .....</b>	<b>12</b>
A.	HHID – HOUSEHOLD IDENTIFIER.....	12
B.	PN – PERSON NUMBER.....	12
C.	INWW – WAVE IDENTIFIER .....	12
<b>VI.</b>	<b>PROGRAM STATEMENTS .....</b>	<b>12</b>
A.	USING THE FILES WITH SAS .....	12
B.	USING THE FILES WITH SPSS .....	13
C.	USING THE FILES WITH STATA .....	13
<b>VII.</b>	<b>OBTAINING THE DATA.....</b>	<b>13</b>
A.	REGISTRATION AND DOWNLOADING THE DATA .....	13
B.	CONDITIONS OF USE .....	13
C.	PUBLICATIONS BASED ON DATA .....	14
<b>VIII.</b>	<b>IF YOU NEED TO KNOW MORE .....</b>	<b>14</b>
A.	HRS INTERNET SITE .....	14
B.	CONTACT INFORMATION.....	14
<b>APPENDIX A: LOADING HRS DATA PRODUCTS IN A MACINTOSH OSX ENVIRONMENT.....</b>		<b>15</b>
	<i>SPSS.....</i>	<i>15</i>
	<i>STATA.....</i>	<i>15</i>

## I. Introduction and Overview

In this document, we outline assumptions, calculations, and other details pertaining to the construction of Social Security wealth measures for “pre-retirement” respondents<sup>1</sup> 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 (NIH U01 AG009740), with supplemental support from the Social Security Administration (SSA). 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 (SS) wealth to respondents in the 1992, 1998, and 2004 waves of the study.

In Version 2 of these data, we improved both the number of cases needing imputations and our imputation strategy. The former improvement was the result of linking more cases that we incorrectly unmatched in Version 1. The latter improvement corrected errors with the former imputation strategy and modified the spouse imputations procedure so that unlinked spouses are imputed only once (for more details see Section III.G).

By receiving these data, which have 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 Section VII for additional details.

## II. Background

The *Prospective Social Security Wealth Measures of Pre-Retirees* data set consists of respondent-level, cross-sectional files constructed from the employment sections of the HRS 1992 (wave 1), HRS 1998 (wave 4), HRS 2004 (wave 7) and the restricted SSA summary and detailed earnings and benefits files. In this public use file, we calculate wealth only for individuals who have not yet retired (as evidenced by claiming SS benefits) (see Section III.C). Analysts interested in SS wealth for current retirees should apply for access to our restricted version of this file.<sup>2</sup> Each individual is uniquely identified by the concatenation of the household ID and the person number, HHID and PN. We organize the data to match the organization of the RAND HRS data files.

## III. Procedures

### A. Initial Calculations and Assumptions

We use the ANYPIA program (Office of the Actuary, v 2008.1) to compute each respondent’s Average Indexed Monthly Earnings (AIME) and Primary Insurance Amount (PIA). We discuss the program’s algorithm and our relevant assumptions in this section.

The ANYPIA program determines eligibility and coverage based on the individual’s earnings record. We have earnings as early as 1951. We project earnings beyond the wave date following Mitchell, Olson and Steinmeier (2000) for individuals who have yet to attain the relevant claim age (62, full retirement age (FRA), and 70) at the wave date. Earnings in the five-

---

<sup>1</sup> By “pre-retirement,” we mean that the respondent has not yet claimed Social Security. In the restricted version of these data, we calculate wealth measures for individuals who have already claimed Social Security based on their current monthly benefits from the SS Master Beneficiary Record.

<sup>2</sup> <http://hrsonline.isr.umich.edu/index.php?p=resproinfo&iyear=1045>

year period preceding the wave year ( $w$ ) are indexed to the  $w - 1$  levels using the National Average Wage Index (AWI) of the SSA based on the Trustees Report for each wave year. The indexed wages are then averaged, with years  $w-1$  through  $w-5$  single year earnings given weights 5, 4, 3, 2, and 1, respectively. Therefore, the projected real earnings for the wave date ( $w$ ) are:

$$earn_w = (1 + CPI_w) * \left( 5 * earn_{w-1} + 4 * \left( earn_{w-2} * \frac{x_{w-2}}{x_{w-1}} \right) + 3 * \left( earn_{w-3} * \frac{x_{w-3}}{x_{w-1}} \right) + 2 * \left( earn_{w-4} * \frac{x_{w-4}}{x_{w-1}} \right) + earn_{w-5} * \frac{x_{w-5}}{x_{w-1}} \right)$$

[i]

where  $earn_w$  represents the SS earnings at time  $w$  and  $x_w$  equals the national wage average at time  $w$ . Then, ANYPIA automatically indexes this value ( $earn_w$ ) upward each year until the claim year, using the National AWI from the wave date Trustee Report.

In cases where a respondent has \$0 earnings in year  $w-1$ , we examine his self-reported work status at the wave date,  $w$ . If he reports that he is unemployed in  $w$ , we project a wage of \$0 beyond  $w$ . However, if he reports that he is employed in  $w$ , we use the weighted average method as just described. In cases where we have the respondent's earning history, but not up until the wave date, then we project his earnings forward from the last observed earnings following the ANYPIA calculator assumptions.

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. The PIA is the monthly amount payable to the worker in claim year dollars. We adjust this in the present value calculations to wave date dollars (see discussion below). It is reduced for claim dates prior to full retirement age and credited for claim dates after full retirement age. We calculate annual PIAs as of three claim ages—age 62, FRA, and age 70. Note that we use the adjusted PIA conditional on the age claim.

Once monthly PIA benefits at each claim date are calculated, we merge those benefits to HRS data files (including the RAND fat file) for the relevant wave date in order to obtain other relevant variables (marital status, education, etc.). We calculate present values for each of the relevant claim ages reported in wave date dollars. For individuals who have already started claiming SS benefits as of the wave date, we calculate present values based on their actual claims (from the benefits file). We use the SSA life tables by year of birth and sex to calculate survival probabilities. We use the following interest rates from the SSA Trustee Reports as of the wave year:

Wave date	Nominal Interest	Inflation	Real Interest
1992	6.3	4	2.2
1998	6.3	3.5	2.7
2004	5.8	2.8	2.9

## B. Wealth for Respondents Who Have Not Yet Claimed

### 1. Background: Types of Benefits

We focus on three distinct types of Social Security benefits in this dataset. A retired worker age 62 or over who is fully insured is eligible to receive monthly cash **retirement insurance benefits** (also called old-age insurance benefits) that are based on his own lifetime earnings (see §300 of the SSA Handbook). In addition, some individuals may be eligible for other auxiliary benefits, such as a monthly spouse’s insurance benefit. This **incremental auxiliary spouse benefit** is based on the spouse’s lifetime earnings (see §305 of the SSA Handbook). In this case, the fully insured retired worker receives a monthly benefit that is comprised of both the “retirement insurance benefit” and the “auxiliary spouse benefit.” Finally, individuals may be eligible to receive a monthly **incremental survivor benefit**, which is based on the deceased insured worker’s lifetime earnings (see §400 of the SSA Handbook). In this dataset, we calculate survivor benefits for widow(er)s only (see §401-409 of the SSA Handbook). We calculate wealth based on each of these monthly benefits separately. In this way, we parse the monthly benefit into 1) the retirement insurance benefit and 2) the incremental auxiliary spouse benefit in cases where both spouses are alive. We parse the monthly benefit into 1) the retirement insurance benefit and 2) the incremental survivor benefit in cases where one worker is deceased.

We offer the following examples (borrowed from Panis et al., 2000).<sup>3</sup>

**Example 1.** Mary and Paul are married. Mary gets \$200 a month in SSA retirement based on her own work history. Paul gets \$1000 a month for his SSA retirement insurance benefits. Mary is entitled to an auxiliary spouse benefit in the amount of \$500 from Paul’s account. Since the \$500 is greater than her own retirement insurance benefit amount (\$200), she is *dually entitled*. She will receive \$200 against her own account and an additional \$300 against Paul’s. We refer to the \$200 portion of this benefit as her “retirement insurance benefit” and the \$300 portion as her “incremental auxiliary spouse benefit.” Note that she actually receives this all in one check in the amount of \$500. If Mary’s own benefit were \$600 instead of \$200, she would not be termed dually entitled and would have no auxiliary spouse benefit. Similarly, Paul is not dually entitled on Mary’s account because he is only eligible for half her benefit (\$100), which is less than what he receives on his own account. Therefore, for the construction of the wealth measures in this dataset, here is a recapitulation of how we define the types of benefits in this case:

	<u>Retirement Insurance benefit</u>	<u>Incremental Auxiliary Spouse benefit</u>
Mary	\$200	\$300
Paul	\$1000	\$0

**Example 2.** Suppose Paul passes away. As Paul’s widow, Mary is entitled to the full amount of his entitlement, \$1000. Mary would receive \$200 based on her own lifetime earnings and \$800 from Paul’s account, for a total monthly benefit in the amount of

---

<sup>3</sup> Panis, C., Euller, R., Grant, C. Bradley, M. Peterson, C.E., Hirscher, R, and Steinberg P. (2000). SSA Program Data User’s Manual. Produced by RAND. PM-973-SSA.

\$1000. We refer to the \$200 portion again as her “retirement insurance benefit” and the \$800 as her “incremental survivor benefit.” Now, the data would change:

	<u>Retirement Insurance benefit</u>	<u>Incremental Auxiliary Spouse benefit</u>	<u>Incremental Survivor benefit</u>
Mary	\$200	\$0	\$800
Paul	\$0	\$0	\$0

We note these differences as we calculate the PV of these three separate streams of income each adjusted by the appropriate survival probabilities.

## 2. Wealth Based on Retirement Insurance Benefits

Using the SS mortality tables<sup>4</sup> and the relevant interest and inflation rates as projected from the wave date Trustee Reports, we calculate the present value of the stream of Social Security retirement insurance benefits based on the retirement insurance benefit for each claim date. Note that for individuals who would also be receiving an incremental auxiliary spouse benefit, wealth based on own earnings *excludes* the portion of the monthly benefit (or PIA) attributable to the spouse’s lifetime earnings (i.e. an auxiliary spouse or survivor benefit). The latter are included in the household wealth measures (see Section III.B.3) and in the restricted version of these data. The expected discounted present value of benefits is given by:

$$PV_c = PIA_c \sum_{t=0}^T P_t *(1+r)^{-t} \quad [a]$$

where

$PIA_c$  = the annual SS Primary Insurance Amount at the projected claim date ( $c$  = age 62, FRA, or age70) reduced for early retirement or credited for delayed retirement and in claim year dollars

$P_t$  = the probability of being alive at time  $t$  conditional on having survived through the wave date

$r$  = real interest rate =  $(1 + \text{nominal inflation rate}) / (1 + \text{inflation rate}) - 1$

We sum discounted benefits from the claim year ( $t=0$ ) until the age of 119. The first term represents the PIA conditional on claim age of  $c$ . We adjust this value by the probability of survival through time  $t$  and discount it back to claim date,  $c$ , dollars. We use yearly compounding since our mortality data are at yearly intervals. To make these values comparable across individuals, we then move  $PV_c$  from claim year dollars to wave date dollars—this is what is reported in the data files. Also, note that we adjust the first year of benefits to account for the fractional consumption of PIA in that year. For example, if an individual is eligible in June, then we use one-half of the relevant PIA 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 2004 will have missing values for  $r7sswr$  (wealth based on own earnings record and early retirement claim age) and  $r7sswrnr$  (wealth based on own earnings record and full retirement claim age).

<sup>4</sup> We use expected survival probabilities based on gender and birth cohort up to age 120.

### 3. Total Household Wealth

In the restricted version of this data, we calculate wealth in three separate components based on the retirement insurance benefit (as described in Section III.B.1), the incremental auxiliary spouse benefit and the incremental survivor benefit. Each period, in addition to any adjustments made for early or late claiming, we weight these components by the appropriate survival probabilities. To make this more concrete, let  $i$  and  $j$  represent the spouses, respectively. Spouse  $i$ 's total Social Security wealth is defined as (assuming adjustments for early/late claiming have already been made to PIAs):

$$\begin{aligned}
 SS\ wealth_i &= PIA_i \sum_{t=0}^T P_{it}(1+r)^{-t} + \\
 &max(0.5PIA_j - PIA_i, 0) \sum_{t=0}^T P_{it}P_{jt}(1+r)^{-t} + \\
 &max(PIA_j - PIA_i, 0) \sum_{t=0}^T P_{it}(1 - P_{jt})(1+r)^{-t}
 \end{aligned}
 \tag{b}$$

where

$PIA_k$  = the annual SS Primary Insurance Amount for spouse  $k$  at the projected claim date reduced for early retirement or credited for delayed retirement and in claim year dollars  
 $P_{ki}$  = the probability spouse  $k$  is alive at time  $t$  conditional on having survived through the wave date

$r$  = real interest rate =  $(1 + \text{nominal inflation rate}) / (1 + \text{inflation rate}) - 1$

The first term represents SS wealth based on the retirement insurance benefit; the second term represents SS 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 SS wealth based on the incremental survivor (widow or widower) benefit, also adjusted survival probabilities. Note that incremental auxiliary spouse and survivor wealth are zero for the higher earner (by definition). To make these values comparable across individuals, household wealth is reported in wave date dollars.

Therefore, total household wealth is defined as:

$$HH\ SS\ wealth_{ij} = SS\ Wealth_i + SS\ Wealth_j
 \tag{c}$$

We provide more details about the construction of wealth based on the incremental auxiliary spouse and survivor benefits in the documentation for the restricted version of these data.

### C. Wealth for Respondents Who Have Claimed

Individuals are designated as already having claimed if they have started receiving benefits as of the relevant wave date according to the Social Security Master Beneficiary Record file (MBR). In this version of the data, we flag these individuals using `rWclaimed` and `sWclaimed`. They will have missing values for all the SS wealth prediction variables. They are not included in the calculation of household wealth. That is, in a household where one member has claimed while

the other one has not, the household wealth only include the wealth of the member who has not claimed.

#### D. Variable Naming Conventions

The variable names are consistent with the pattern used in the RAND HRS data. The first character indicates whether the variable refers to the respondent, “r,” the spouse, “s” or the household, “h.” The second character corresponds to the wave to which the variable pertains: “1” for 1992, “4” for 1998, and “7” for 2004. All Social Security wealth measures are then followed by “ssw.” The sixth character denotes the source of the benefit: “r” for retirement insurance benefit. The seventh and eighth 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.

#### E. Variables for Each Wave

For each wave, we calculate benefits assuming claim ages that the respondent has not yet attained as of the wave date, otherwise, we assign missing values if the respondent is older than a particular claim age. We also assign missing values if the respondent has already started claiming benefits as evidenced by being in the Social Security MBR file. The following table shows which wealth values we calculate for individuals of a certain age as of the wave date:

*Assuming claim age of:*

	Age 62	FRA	Age 70
70+	n/a	n/a	n/a
68-69	n/a	n/a	✓
65-67	n/a	✓	✓
63-64	n/a	✓	✓
62 or younger	✓	✓	✓

*Age as of wave date*

#### F. Data Codebook

VARIABLES	NAME	BRIEF DESCRIPTION
Respondent retirement insurance benefit wealth, assuming an age 62 claim	rWsswrer	Respondent’s predicted SS wealth based on the retirement insurance benefit (calculated based on own earnings records) assuming a claim age of 62.
Respondent retirement insurance benefit wealth, assuming a full retirement age claim	rWsswrnr	Respondent’s predicted SS wealth based on the retirement insurance benefit (calculated based on own earnings records) assuming a full retirement claim age.
Respondent retirement insurance benefit wealth, assuming an age 70 claim	rWsswrxa	Respondent’s predicted SS wealth based on the retirement insurance benefit (calculated based on own earnings records) assuming a claim age of 70.
Spouse retirement	sWsswrer	Spouse’s predicted SS wealth based on the retirement



insurance benefit wealth, assuming an age 62 claim		insurance benefit (calculated based on own earnings records) assuming a claim age of 62.
Spouse retirement insurance benefit wealth, assuming a full retirement age claim	sWsswrnr	Spouse's predicted SS wealth based on the retirement insurance benefit (calculated based on own earnings records) assuming a full retirement claim age.
Spouse retirement insurance benefit wealth, assuming an age 70 claim	sWsswrx	Spouse's predicted SS wealth based on the retirement insurance benefit (calculated based on own earnings records) assuming a claim age of 70.
Household total SS wealth, assuming all claim at age 62	hWsswr	The sum of respondent's retirement insurance wealth, respondent's incremental auxiliary spouse benefit wealth, respondent's incremental survivor benefit wealth, spouse's incremental retirement insurance benefit wealth, spouse's incremental auxiliary spouse benefit wealth, and spouse's incremental survivor benefit wealth all assuming both respondent and spouse claim at age 62, but excluding wealth of respondents who had already claimed as of the time of interview. (See Section III.B.3 for more details.)
Household total SS wealth, assuming all claim at full retirement age	hWsswrnr	Same summation as above, but assuming both respondent and spouse claim at their respective full retirement ages.
Household total SS wealth, assuming all claim at age 70	hWsswrx	Same summation as above, but assuming both respondent and spouse claim at age 70.
Respondent currently claiming benefits flag	rWclaimed	An indicator variable equal to one if the respondent has already claimed SS benefits as of the wave date, W. If rWclaimed = 1, then all respondent retirement insurance benefit wealth variables will be missing.
Spouse currently claiming benefits flag	sWclaimed	An indicator variable equal to one if the spouse has already claimed SS benefits as of the wave date, W. If sWclaimed = 1, then all spouse retirement insurance benefit wealth variables will be missing.
Wave status: Response indicator (from RAND)	inwW	=1 if the respondent responded in wave W

W refers to the wave date: W=1 refers to the 1992 wave date, W=4 refers to the 1998 wave date, and W=7 refers to the 2004 wave date.

### 1. Rounding and Top-Coding

Wealth variables are first rounded to the nearest \$100 and then top-coded at the top 2 percent. PIAs are rounded to the nearest \$5 and then top-coded at the top 2 percent. Values exceeding the top 2% have been replaced with the mean of the top 2% for each wave.

## G. Imputations

We impute own and spouse PIA in cases where the respondent and/or the respondent’s spouse are unmatched to Social Security earnings records. In cases where we have *some* earnings records for an individual, but not all, we do not impute wealth. For example, if an individual is linked in 1992, but not in 2004, we will not impute wealth for him for any wave. In the 1998 and 2004 waves (where we would be missing earnings records from 1992 onwards), we simply extrapolate earnings from 1992 following the assumptions as discussed in Section III.A. We define imputation strategies in the following sections.<sup>5</sup>

### 1. Respondent Imputations

If the respondent is unmatched and expects to claim SS (self-reported), we impute the respondent’s own PIA using a mix method: regression based nearest neighbor hot-decking. Using the sample of matched respondents, we regress the respondent’s own PIA on several demographic and labor force characteristics.<sup>6</sup> We then sort all observations based on their predicted PIA from the estimation. We substitute the missing respondent’s PIA with an actual PIA from the nearest matched respondent (hot-deck). In these cases, we assume the respondent is insured or will be at the time of claim.

In cases, where the respondent is unmatched and does not expect to claim SS or does not report expectations we impute both the respondent’s insured status and own PIA. We estimate a 2-step Heckman model to correct for sample selection. We use an indicator variable as the exclusion restriction in the selection equation: whether the individuals ever worked as a federal employee.<sup>7</sup> Civilian federal employees hired before 1984 are excluded from OASDI. From this estimation, we hot-deck both insured status and PIA from matched nearest neighbors.

#### a) By Wave Year

**1992<sup>8</sup>**

<i>r1impute</i>	<i>Freq.</i>	<i>Percent</i>	<i>C. Freq.</i>
1	10,212	80.72	80.72
2	1,933	15.28	96.00
3	506	4.00	100.00
Total	12,651	100	

**1998<sup>9</sup>**

<i>r4impute</i>	<i>Freq.</i>	<i>Percent</i>	<i>C. Freq.</i>
1	16,845	73.63	73.63
2	1,578	6.90	80.52
3	4,456	19.48	100.00
Total	22,879	100	

<sup>5</sup> The imputation flags, *rWimpute* and *sWimpute*, are only available in the restricted version of this data.

<sup>6</sup> We include the following explanatory variables: respondent’s age, three marital status indicators, two education indicators, an indicator for whether the respondent was born outside of the US, three labor force status indicators, five total work experience indicators, a union status indicator, three pension plan indicators, a health insurance indicator, a home owner indicator, annual wages for current or last job and previous job, total household income, eight region indicators, and occupation and industry indicators for current or last job and previous job.

<sup>7</sup> *Ibid.*

<sup>8</sup> One observation is missing because we do not have a value for birth year.

<sup>9</sup> Two observations are missing because we do not have values for birth year.

**2004<sup>10</sup>**

<i>r7impute</i>	<i>Freq.</i>	<i>Percent</i>	<i>C. Freq.</i>
1	15,945	73.08	73.08
2	1,989	9.12	82.19
3	3,886	17.81	100.00
Total	21,820	100	

## 2. Spouse Imputations

In cases where the spouse is unmatched to the SS data and the spouse was imputed as a respondent as described in Section III.G.1, we impute the spouse’s PIA as described in the previous section, Section III.G.1.

In cases, where the spouse is unmatched to the SS data and the spouse is not imputed as a respondent as described in Section III.G.1, we impute the spouse’s PIA using only characteristics of the *respondent* for the purpose of calculating a spousal or widow benefit for the respondent only. That is, we do not actually impute any wealth (own, spouse or widow) to the missing spouse. We use the imputed PIA to calculate potential auxiliary benefits to the *respondent*. In these cases, the spouse is missing—s/he was not interviewed during that particular wave. Using the sample of matched respondents, we regress the spouse’s PIA on demographic and labor force characteristics of the *respondent* (since we do not have this information on the spouse).<sup>11</sup> We then sort all observations based on their predicted spouse PIA from this estimation. We substitute the missing spouse’s PIA with an actual spouse PIA from the nearest matched respondent (hot-deck).

---

<sup>10</sup> One observation is missing because we do not have a value for birth year.

<sup>11</sup> We include the following explanatory variables: respondent’s age and its square, respondent’s gender, two race indicator variables for the respondent, two labor force status indicator variables for the respondent, respondent’s education and its square, respondent’s labor market experience and its square, respondent’s annual earnings and its square, respondent’s experience interacted separately with respondent’s gender, earnings, and age, respondent’s age interacted separately with respondent’s earnings and experience, total household income, a house ownership indicator variable, indicator variables for whether the respondent is currently claiming SS or expects to claim in the future, the length of the current marriage, two indicator variables for whether the respondent is divorced or widowed, and eight region dummy variables,

a) *By Wave Year*

**1992<sup>12</sup>**

<i>s1impute</i>	<i>Freq.</i>	<i>Percent</i>	<i>C. Freq.</i>
1	7,808	78.93	78.93
2	1,724	17.43	96.36
3	360	3.64	100.00
Total	9,892	100	

**1998<sup>13</sup>**

<i>s4impute</i>	<i>Freq.</i>	<i>Percent</i>	<i>C. Freq.</i>
1	10,493	75.14	75.14
2	3,020	21.63	96.77
3	451	3.23	100.00
Total	13,964	100	

**2004<sup>14</sup>**

<i>s7impute</i>	<i>Freq.</i>	<i>Percent</i>	<i>C. Freq.</i>
1	9,290	73.14	73.14
2	3,078	24.23	97.37
3	334	2.63	100.00
Total	12,702	100	

#### IV. Distribution Information

The *Prospective Social Security Wealth Measures of Pre-Retirees* data are distributed in 1 data file at the respondent level with three years of SS wealth measures: 1992, 1998, and 2004. The file consists of 29,070 person-year observations, 44 variables and the primary identifiers (HHID, PN). The records in the data files are sorted in order by these primary identifiers. The data are provided in ASCII format, with fixed-length records. Use associated SAS, SPSS, or STATA program statements to read the data into the analysis package of your choice. The file is packaged for download from our website in ZIP format.

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

- .DA for data files,
- .SAS for SAS program statements,
- .SPS for SPSS program statements,
- .DO for Stata DO statements,
- .DCT for Stata dictionary statements, and
- .TXT for codebook files.

For example,

<sup>12</sup> One observation is missing because we do not have a value for birth year.

<sup>13</sup> Two observations are missing because we do not have values for birth year.

<sup>14</sup> One observation is missing because we do not have a value for birth year.

SSWEALTHP.DA contains ASCII data.  
SSWEALTHP.SAS contains corresponding SAS program statements,  
SSWEALTHP.SPS contains corresponding SPSS program statements,  
SSWEALTHP.DO contains corresponding Stata DO statements,  
SSWEALTHP.DCT contains corresponding Stata dictionary statements, and  
SSWEALTHP.TXT contains the ASCII codebook.

The Social Security Wealth 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).

## V. Identification Variables

### A. 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.

### B. 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.

### C. inwW – Wave Identifier

In combination with HHID and PN, inwW uniquely identifies each person-year observation.  $W = 1, 4,$  and  $7$  for the 1992, 1998, and 2004 waves, respectively.

## VI. 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 a .SPS, those with SAS statements with a .SAS extension, and those with Stata statements with .DO and .DCT extensions. The following examples assume a Microsoft Windows Environment. For information on working within a Macintosh OSX environment, see Appendix A.

### A. Using the Files with SAS

To create a SAS system file from the ASCII data file, load the .sas program statement files into the SAS Program Editor and reference the appropriate .da data files. If the \*.sas file is located in "c:\wealthp\sas" and the data file is located in "c:\wealthp\data", you can run the file as is. A SAS system file (\*.sas7bdat, assuming that you are using SAS V8.0 or higher) will be saved to directory "c:\wealthp\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.

## B. Using the Files with SPSS

To create an SPSS system file from the ASCII data file, load the .sps program statement files into the SPSS syntax editor window, reference the appropriate .da data files, and select the Run>All option. If the \*.sps file is located in "c:\wealthp\spss" and the data file is located in "c:\wealthp\data", you can run the file as is. An SPSS system file (\*.sav) will be saved to directory "c:\wealthp\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.

## C. Using the Files with Stata

To use Stata with the ASCII data file, three file types must be present for that data set: .dct, .do, and .da. 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:\wealthp\stata" and the data file is located in "c:\wealthp\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.<sup>15</sup>

# VII. Obtaining the Data

## A. Registration and Downloading the Data

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

## B. 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

---

<sup>15</sup> 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. You may do this by reading the .DCT file into a text or word processing program and changing the case. For instance in Microsoft Word, Edit, Select All, Format, Change Case, lowercase.

- To report apparent errors in the HRS data and documentation files
- To notify us (via our website) of changes in your contact information

For more information concerning privacy issues and conditions of use, please read *Conditions of Use for Public Data Files* and *Privacy and Security Notice* at the Public File Download Area of the HRS Web site.

### **C. 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 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  
The Institute for Social Research, Room 3050  
P.O. Box 1248  
Ann Arbor, MI (USA) 48106-1248

## **VIII. 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 Web site, or if you have any comments, please contact us. We will do our best to provide answers.

### **A. HRS Internet Site**

Health and Retirement Study public release data and additional information about the study are available on the Internet. To access the data and other relevant information, visit the HRS Web site: <https://hrsonline.isr.umich.edu/>

### **B. Contact Information**

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

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

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

## Appendix A: Loading HRS Data Products in a Macintosh OSX Environment.

Users must modify the Windows file structure syntax to match that of OSX. Open the SPSS syntax file or the Stata do/dct files in an ASCII editor such as TextEdit and make the changes indicated below.

### SPSS

In this example, we assume that the user has downloaded the **2011 Internet** study and placed the files in a **Desktop** folder called **2011Inet** with the ASCII data file stored in subfolder **data** and the syntax file in subfolder **spss**. Then the commands in the syntax file would be modified to look like this:

```
FILE HANDLE net11_R /name='Desktop/2011Inet/data/Net11_R.sav' LRECL=1812.
DATA LIST FILE= net11_R/
HHID 1-6(A)
[rest of syntax file goes here]
.
execute.
SAVE /outfile 'Desktop/2011Inet/spss/Net11_R.sav'.
Execute.
```

### STATA

In the following example we assume that:

- The username is “user1”
- The zip file containing tracker information has been downloaded to the user’s desktop from the HRS file download site
- The user has decompressed the zip file (use Stuffit) into a desktop folder named trk2012
- The statistical package is stata

File TRK2012TR_R.do should be modified as follows	
<i>Change PC version from...</i>	<i>To OSX version...</i>
infile using c:\trk2012\stata\TRK2012TR R.dct	infile using /Users/user1/Desktop/trk2012/stata/TRK2012TR R.dct
save c:\trk2012\stata\TRK2012TR R.dta	save /Users/user1/Desktop/trk2012/stata/ TRK2012TR R.dta
File TRK2012TR_R.dct should be modified as follows	
<i>Change PC version from...</i>	<i>To OSX version...</i>
dictionary using c:\trk2012\data\TRK2012TR R.da {	dictionary using /Users/user1/Desktop/trk2012/data/TRK2012TR R.da {