

Health and Retirement Study

**Imputations for Employer-Sponsored  
Pension Wealth from Current Jobs in 2004**

Version 1  
July 2009

Data Description and Usage

# 1. Overview and Background

The *Imputations for Employer-Sponsored Pension Wealth from Current Jobs in 2004* (Version 1.0) 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, cross-sectional data for researchers and policymakers who study aging. Funding for the Health and Retirement Study is provided by the National Institute on Aging at NIH (U01 AG009740), with supplemental support from the Social Security Administration. The study is conducted by the Institute for Social Research (ISR) at the University of Michigan.

The Health and Retirement Study 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.<sup>1</sup> 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. In 2004 a new sample cohort (Early Baby Boomer -- EBB) of individuals born in 1948-53 (age 51-56 in 2004) was introduced. Future plans call for a new 6-year birth cohort to be enrolled every six years, beginning in 2010. All cohorts are drawn as nationally representative samples, with those based on the 1992 screen containing over-samples of African-Americans and Hispanic Americans.

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 the respondents. In addition, you agree to send HRS a copy of any publications you produce based on the data. See [Obtaining the Data](#) for additional details.

## 2. Procedures

### 2a. Introduction

Pension wealth estimates can be constructed either from a reported current account balance, or from calculating the present discounted value of a stream of future benefits. For this version of pension wealth imputations, we follow previous practice for 1992 and 1998 in using the respondent's self-reported pension plan type to determine which method of wealth estimation to use. If a respondent reported being covered by least one DC plan on his current job in 2004, his/her DC wealth for each such plan is taken from the 2004 survey self-reported account balances. If a respondent reported being covered by at least one DB plan in 2004, the DB

---

<sup>1</sup> 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, Medicaid Services, or CMS).

pension wealth value for each such plan is calculated by the HRS Pension Estimation Program using a combination of self-reported data from the employment section (J) of the 2004 survey and pension plan rules obtained from the plan's Summary Plan Description (SPD). If a respondent reported being covered by plans of both types or a Combination plan, both DB and DC wealth numbers are calculated.

Imputations are necessary when a) a respondent fails to provide an account balance for a self-reported DC plan, or b) no SPD is available for the respondent's self-reported DB pension plan.

SPDs used in this version of DB pension wealth were obtained from one of two sources:

- Employers responding to the Employer Pension Provider Survey (EPPS) conducted in 2005.
- Respondents who obtained an SPD from their employer in response to a request made during the 2004 and 2006 HRS surveys.

## **2b. Pension Wealth from Defined Benefit Plans**

*Pension Estimation Program* is used to calculate the present values from Defined Benefit (DB) plans for respondents linked to a pension plan in 2004. If that link is missing, the imputations for pension wealth values were performed. (see *Imputation Methods*).

DB values are calculated or imputed for each respondent at seven different ages. Those include the expected age of retirement reported by respondents with DB or combination plans, early age of retirement, normal age of retirement, ages 60, 62, 65 and 70. For respondents linked to pension plans, normal and early retirement ages are derived from pension plan provisions. For respondents with missing linkage, these ages are imputed.

The calculations use the intermediate future real interest rate (2.9%) and the inflation rate (2.8%) forecast by the Social Security Administration.

### **Present Value Calculations**

General formula used for computing present values of retirement benefits:

$$PV(QuitDate) = \left( \sum_{t=0}^{119-age(QuitDate)} P(\text{alive at } t) * \left[ \frac{1+COLA}{1+i_{nom}} \right]^t * B(QuitDate, t) \right)$$

$$i_{nom} = (1 + Inflation Rate / 100)(1 + Real Interest Rate / 100) - 1$$

where :

$$P(\text{alive at } t) = \frac{NumberAlive(QuitDate + t, Sex, BirthCohort)}{NumberAlive(QuitDate, Sex, BirthCohort)}$$

probability of surviving for  $t$  years after quitting conditional on being alive at quit date, computed using sex- and birth cohort-specific mortality tables.

$NumberAlive$  = the number of people still living at a given date assuming a population of 100,000 at birth.

$COLA$  = plan-specific annual growth rate of payment

$i_{nom}$  = nominal interest rate

$B(QuitDate, t)$  = pension benefit in year  $t$  after the quit date

The present values obtained by this formula are then discounted back to 2004.

## **2c. Pension Wealth from Defined Contribution Plans**

To construct pension wealth from Defined Contribution (DC) plans, we used self-reported account balances from current job. A respondent could report multiple accounts balances from the same job. If some of these amounts were missing, they were imputed (see *Imputation Methods*). The imputations were conducted for each reported account separately.

The total DC wealth is computed as a sum of all accounts from current job in 2004.

## **2d. Imputation Methods**

Three methods of imputation were used: mixed, hot-decking, and replacement. In the mixed method, we use a regression to get the predicted value of the variable to be imputed. Then the data are sorted by the predicted variable. At the last step, the missing/DK/RF values are replaced with the observed value of its preceding adjacent observation. This method is used when the sample size is large enough to run a regression with firm size, industry, union membership, and personal characteristics as explanatory variables to estimate the predicted values.

In the hot-decking method, we generate a set of random numbers and sort the data by the random values. Then the missing/DK/RF value was replaced with the observed value of its preceding adjacent observation.

The third method, the replacement method, is used when the sample size is too small for either the mixed or hot-decking methods. When only one observation has a non-missing value, all missing are replaced with that value.

The codebook that is distributed with these data contains descriptions of methods used in imputations of each variable.

## 3 Distribution Information

### 3a. Distribution Set

The *Imputations for Employer-Sponsored Pension Wealth from Current Jobs in 2004* files are packaged for distribution in a .ZIP file, *ImpPenW2004.zip*, which contains seven files:

- *pwi04.da* ASCII data file
- *pwi04.sas* SAS data descriptors
- *pwi04.sps* SPSS data descriptors
- *pwi04.dct* Stata data descriptors
- *pwi04.do* Stata "do" file
- *pwi04.txt* codebook information
- *ImpPenW2004.pdf* data description (this file)

### 3b. Data File

The data file is a respondent level file containing one record for every respondent who was a part of HRS 2004. Identification variables – HHID and PN are provided for merges with other HRS data files.

Variable *DBFlag04* indicates that a respondent reported being covered by least one DB/Combination plan on his current job in 2004 and hence his/her DB wealth number is constructed (either calculated by the software or imputed).

Variable *DCFlag04* indicates that a respondent reported being covered by at least one DC/Combination plan on his current job in 2004 and hence his/her DC wealth number is constructed (either from self-reported or imputed account balances).

### 3b1. DB Account Balances for 2004

#### *Naming Conventions for DB Variables*

Values	DB60_04x	DB values at age 60
	DB62_04x	DB values at age 62
	DB65_04x	DB values at age 65
	DB70_04x	DB values at age 70
	DBer_04x	DB values at early retirement age
	DBnr_04x	DB values at normal retirement age
	DBxp_04x	DB values at expected retirement age
Imputation flag	DB60_04t	Flag indicating that DB value at age 60 is imputed
	DB62_04t	Flag indicating that DB value at age 62 is imputed
	DB65_04t	Flag indicating that DB value at age 65 is imputed
	DB70_04t	Flag indicating that DB value at age 70 is imputed
	DBer_04t	Flag indicating that DB value at early retirement age is imputed
	DBnr_04t	Flag indicating that DB value at normal retirement age is imputed
	DBxp_04t	Flag indicating that DB value at expected retirement age is imputed

**Note that these values represent the present discounted value of future benefits as of survey year (2004).**

### *Naming Conventions for various retirement ages*

Values	EAGE_04X	EARLY RETIREMENT AGE
	NAGE_04X	NORMAL RETIREMENT AGE
	XAGE_04X	EXPECTED AGE OF RETIREMENT
Imputation flag	EAGE_04T	FLAG - INDICATES THAT EARLY AGE OF RETIREMENT IS IMPUTED
	NAGE_04T	FLAG - INDICATES THAT NORMAL AGE OF RETIREMENT IS IMPUTED
	XAGE_04T	FLAG - INDICATES THAT EXPECTED AGE OF RETIREMENT IS IMPUTED

## **3b2. DC Account Balances for 2004**

### *Naming Conventions for DC Variables*

In Section J of 2004 HRS, there are 9 questions about account balances. The imputations are performed for each account separately. In the imputation file, the names of the original HRS variables are preserved (with the letter “X” added to the variable name to indicate that the missing values are imputed). For example, variables *JJ339AX*, *JJ339BX*, *JJ339CX* and *JJ339DX* contain imputed or reported account balances. The corresponding set of variables with letter “T” at the end, *JJ339AT*, *JJ339BT*, *JJ339CT* and *JJ339DT*, indicate type of imputations performed for the missing values. Summary variable *DC\_04* indicates the values from all DC or combination accounts reported.

## **3b3. Calculating Total Pension Wealth**

In order to compute total pension wealth, a user needs to add DB values at a given age of retirement (60, 62, 65, 70, early, normal or expected) and the values of all DC/Combination accounts in 2004 *DC\_04*.

## **3c. Directory Structure**

Although a particular setup is not required for using HRS files, we have traditionally suggested a directory structure. By using this directory structure, you will not have to change the path name in your data descriptor files. If you use a different structure, just change the directory references in the program statement files.

Directory	Contents
=====	=====
c:\pwimp04	File downloaded from Web site
c:\pwimp04\codebook	codebook and documentation files
c:\pwimp04\data	data files
c:\pwimp04\sas	SAS command files
c:\pwimp04\spss	SPSS command files
c:\pwimp04\stata	Stata command files

Decompress the .zip file into the appropriate subdirectories. You will need approximately 10 MB of free space on your storage device to store these files.

## **4. Identification Variables**

### ***4a. HHID – Household Identifier***

In the initial wave of data collection (in 1992 for the HRS sub-sample, in 1993 for the AHEAD sub-sample, and in 1998 for the WB and CODA sub-samples) each sample household was assigned a Household Identifier. HHID 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. HHID has six digits.

### ***4b. 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. Program Statements**

The Employer Pension Tracker File comes with associated SPSS, SAS, or Stata program statements to read the data. Files containing SPSS statements are named with a .SPS extension, those with SAS statements with a .SAS extension, and those with Stata statements with .DO and .DCT extensions.

### ***5a. Using the Files with SAS***

To create a SAS system file for a particular data set, two file types must be present for that data set -- .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:\pwimp04\sas" and the data file is located in "c:\pwimp04\data", you can run the file as is. A SAS system file (\*.SD2 or \*.SAS7BDAT) will be saved to directory "c:\pwimp04\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.

### ***5b. Using the Files with SPSS***

To create an SPSS system file for a particular data set, two file types must be present for that data set -- .SPS program statement files and .DA data files. To create an SPSS system file, you must first open the \*.SPS file in SPSS as an SPSS Syntax File.

If the \*.SPS file is located in "c:\pwimp04\spss" and the data file is located in "c:\pwimp04\data", you can run the file as is. An SPSS system file (\*.SAV) will be saved to directory "c:\pwimp04\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.

### **5c. Using the Files with Stata**

To use Stata with a particular data set, the following three file types must be present for that data set -- .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:\pwimp04\stata" and the data file is located in "c:\pwimp04\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 lowercase variable names, you may wish to convert the .DCT files to lowercase 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.

## **6. Obtaining the Data**

### **6a. Registration and Downloading the Data**

HRS data are available for free to researchers and analysts at the HRS Web site. In order to obtain public release data, you must first register at our Web site. 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 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.

### **6b. Conditions of Use**

By registering, you agree to the Conditions of Use governing access to Health and Retirement 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 or documentation files
- to notify us ([via our Web site](#)) 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* in the Public File Download Area of the HRS Web site.

### **6c. 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, Room 3050  
P.O. Box 1248  
Ann Arbor, MI (USA) 48106-1248

Alternately, you may contact us by e-mail at [hqsquest@isr.umich.edu](mailto:hqsquest@isr.umich.edu) with “Attn: Papers and Publications” in the subject line.

## **7. If You Need to Know More**

This document is intended to serve as a brief overview and to provide guidelines for using the *Imputations for Pension Wealth* data release. 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.

### **7a. 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, point your Web browser to the HRS Web site: <http://hrsonline.isr.umich.edu/>

### **7b. Contact Information**

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

Internet: [Help Desk](#) at our Web site

E-mail: [hqsquest@isr.umich.edu](mailto:hqsquest@isr.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

FAX: (734) 647-1186