

# Health and Retirement Study

Imputations for Pension-Related Variables

Final, Version 1.0

June 2005

Data Description and Usage

# 1. Overview and Background

The *Imputations for Pension-Related Variables* (Final, 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 National Institute on Aging provided funding (NIH U01 AGO9740), with supplemental support from the Social Security Administration. The Institute for Social Research (ISR) Survey Research Center (SRC) at the University of Michigan conducted the survey.

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.

The employment section of the Health and Retirement Study (HRS) focuses on retirement and covers specific information about pension plan(s) that a respondent may have. The pension plans may be from a current job if the respondent is employed at the time of interview or from their last job if s/he was not employed when first interviewed. All respondents are asked about their pension plans, up to three, from previous jobs where the respondent worked five years or more when first interviewed. Pension related questions include a variety of information such as pension coverage, number of plans, type of plan(s), normal and early retirement age, amount of contribution, type and amount of future benefits if the plan is a Defined Benefit (DB), current account balances if the plan is a Defined Contribution (DC) or a combination of DB and DC, and more.

It is believed that often respondents don't have complete information about specifics of their pension plans. In fact the HRS data show a high rate of Don't Know, Refused, and Missing responses regarding plan type and other pension characteristics. In this document we describe the imputation process for the Don't Know, Refused, and Missing responses for key pension related variables from the HRS interview years 1992 (Wave 1), 1994 (Wave 2), 1996 (Wave 3), 1998 (Wave 4), 2000 (Wave 5) and 2002 (Wave 6). The imputed variables are from the employment sections of Wave 1 to Wave 6, which include: sections F, G, and H of Wave 1; FA, FB, FC, G, and H of Wave 2; G, GG, and GH of Wave 3, Wave 4, and Wave 5; J, K, and L of Wave 6. It should be noted that the imputations are performed for the HRS cohort in Waves 1 to 3 and for the HRS and War Baby cohorts in Waves 4, 5 and 6; AHEAD and CODA cohorts are not included in these imputations.

The imputations for each wave are cross-sectional. The information from previous or successive waves is not integrated in a current wave's imputations. For example, for imputing the relevant variables from the 1994 survey, information from the 1992 and 1996-2002 surveys is not used. In the future we will provide another set of imputations that will integrate the information across waves. The imputation process includes determining a control variable(s), which will be defined later, and choosing an appropriate technique for the imputation. However, some of the variables require some type of adjustment or replacement prior to the actual imputation. The adjustment and imputation processes are described below (Section 2).

## 2. Procedures

### 2a. Adjustment Process

Almost all non-Don't Know, non-Refused, and non-Missing values are considered to be valid observed values. But a very small percentage of those observed values may not be valid due to errors in the instrument's code, preload, or some other reasons unknown to us. Therefore, some adjustment should be made to correct for those errors. The adjustment process may involve converting a Missing, Don't Know, or Refused response to an observed value or removing an observed value and setting it to missing or blank prior to any imputations.

Following is a list of major adjustments in Wave 1 to Wave 6:

1. HRS staff imputed some pension related variables in sections G and H of Wave 1. These imputed variables were included in earlier releases. However, some of the imputed values are not consistent with observed values for other pension variables that came earlier in the hierarchy. For example, a value may have been imputed for the disposition<sup>1</sup> of type DB pension plans but there is no reported or imputed value for the pension coverage variable (H12). The imputed values in sections G and H, previously imputed by HRS staff, are not used in this imputation process. We use a different method to impute values when appropriate.
2. Missing values and blanks are recorded as zeros in Wave 2. Although this may not be of any consequence for discrete variables, it is a problem when trying to identify the missing values in continuous variables. For example, an account balance for a DC plan with a zero value may be a zero balance or a missing value. To adjust the Wave 2 ratio of actual zeros to missing values, we have applied the percentage of respondents having a zero balance in Wave 1 to Wave 2. If the percentage of cases with recorded zero in Wave 2 is more than the percentage of cases with recorded zero in Wave 1, we select additional cases randomly and set them to missing until the percentage of cases with recorded zeros is the same in Wave 1 as in Wave 2.
3. Respondents who have changed jobs after their first interview are asked about their pension plan from their previous job (e.g., Wave 1's job in Wave 2, Wave 2's job in Wave 3...). Respondents are asked if they are working at the same employment (G19b) as the previous interview. If the response is "No: not working for the same employer", they are asked when they stopped working for that employer (G26). The follow-up question to G26 is about their pension plan from that job. There is a small percentage of respondents who responded "No: not working for the same employer" in G19b who also responded "still working for the same employer" when asked when they stopped working for that employer in G26. Due to the incorrect answer in G19b, these respondents were asked about their pension plans from their previous job (question G30 and its follow-up questions). This exaggerates the number of cases with a previous pension. The adjustment process for these responses is to set

---

<sup>1</sup> That is H14 in the first pension sequence of H section where the respondent is asked "Do you expect to receive benefits from this plan in the future, are you receiving benefits now, did you get a cash settlement when you left, did you lose your benefits, or what?"

responses for the pension coverage question (G30) and its follow-up questions to blank.

4. In Wave 3, some respondents reported working at the same employment as in Wave 2 and were included in a pension plan in Wave 2 but are skipped over the “same pension”<sup>2</sup> sequence and asked the “new pension” sequence questions. The adjustment for those cases involves transferring the observed values from the pension coverage question in the new pension sequence (E2861) to the first question in the same pension sequence (E2835). If the response is “5. No, don’t have a pension plan” in E2861, the missing value in E2835 is replaced with “7. Denies being covered by a pension”. However, if the response in E2861 is “Yes” then we assign a value of “1. Yes, the rules that govern pension has changed” to E2835. This means that those respondents would have values in the new pension sequence after the first question in the same pension sequence. This is a much simpler process and we can keep all the information collected in the new pension sequence for those respondents. However, the user must note that this assignment may not be completely accurate and would result in exaggerating the number of cases with a response of “1. Yes the rules have changed”. We identify those cases with an R in the accompanying (imputation type) variable, described later.
5. There are some respondents in Wave 3 and Wave 4 with responses for the same pension sequence and should be for the new pension sequence. These are the respondents who now have a job and a pension plan and did not have a job in the previous wave or reported not having a pension plan on the job if working. The adjustment process for those cases is transferring the information from the same pension sequence to the new pension sequence.
6. Respondents are asked about the number of pension plans that they are included in while working at a job. The design of the instrument allows collecting information for up to three plans in Wave 1 through Wave 4 and up to four plans in Wave 5 in the same pension sequence. Up to three of plans in Wave 1 to Wave 4 and up to four

---

<sup>2</sup> Basically, there are two sets of pension questions designed for those respondents who are currently employed in each wave. One set is short and skips some of the questions such as the amount of contribution toward a defined benefit pension plan, or the number of years the respondent has been included in a pension plan, retirement age, etc. This version is designed for those who are working at the same employment as in the previous interview wave and reported that they have been included in a pension plan in that job and the rules that govern their pension plans have not changed since last interview. A respondent who is self-employed currently and has been self-employed in the previous interview and for whom the start date of the business is on or before the last date of the previous interview, also gets asked the shorter sequence. For simplicity, we call this sequence the "same pension" sequence.

A more detailed sequence of pension questions covering a variety of information is designed for those who are currently working for pay who report a pension, and who are interviewed for the first time, or have changed jobs since last interview wave, or did not have a pension plan at the time of last interview if working with the same employment. Those who reported that the rules that govern their pension plan have changed since last interview also go through the larger sequence of questions. We call this version the "new pension" sequence.

If respondents report not included in a pension plan they are asked if the employer offers such plans and the possibility and circumstances that the respondent may become eligible and be included in such pension plans.

plans in Wave 5 are allowed in the new pension sequence. However, in Wave 4 respondents with more than three pension plans in the same pension sequence have been pushed to the new pension sequence after completing the same pension sequence. This has caused confusion for the respondents and will be confusing to the users. We make the adjustment by setting the new pension sequence variables to blank for cases with more than three plans in the same pension sequence in wave 4.

7. In Wave 4, the new spouses from HRS respondents are skipped out of the pension questions from GG (last job) and GH (previous jobs) sections. We assume they have missing values and impute for their pension related variables using the War Baby cohort as donors.
8. There are some other minor adjustments to be made for some of the variables. Since problems seemed not to be systematic and involve only a few cases in the variable involved we will report them in the comment column of Table 3, *Imputations for Pension-Related Variables Cross-Reference*.

## **2b. Imputation Process**

The imputation process involves converting the Don't Know and Refused responses to Missing values (in most cases)<sup>3</sup>, determining control variables, and selecting an appropriate imputation technique based on the sample size, covariates, and information from other related variables. The imputed variables have the names of the original variables name with an extension of "X". In addition to the imputed variables we construct another set of variables indicating the type of imputation. The imputation type variables have an extension of "T" instead of "X".

### **2b1. Control Variables**

Control variables are those variables that help us to determine which cases should have responses for a variable we want to impute. We may need only one variable as the control variable or many more depending on the specifics of the variable to be imputed. For example, for imputing the pension coverage variable for self-employed respondents in Wave 1 only one variable serves as the control variable; "who work for?" (F3: V2718). But for the variable representing G69 in Wave 4 asking if "the rules that govern your pension have changed since the last interview" many more variables are required. In this case we must know if a respondent who is an employee is working at the same employment since Wave 3 and if s/he was included in a pension plan at that time. If the respondent is self-employed we want to know if (s)he was self-employed also in Wave 3 and if the start date of the business was before the Wave 3 interview date and (s)he was included in a pension plan at that time. If the respondent was not interviewed in Wave 3 we have to check that information against the same information provided in Wave 2 and similarly for Wave 1 if not interviewed in Wave 2 either.

---

<sup>3</sup> In some cases the Don't know and Refused responses are replaced by a particular outcome. For example by the design of the instrument, "the type of pension plan" the Don't Know and Refused responses are replaced with type DB. Moreover, for the most part the information collected for a DB plan will apply to a DC plan also.

## 2b2. Imputation Methods

The imputation techniques used in this project include Mixed Method, Hot-decking, and Replacement. The appropriateness of the technique for each variable depends on the sample size, availability of covariates<sup>4</sup>, and information from other related variables.

### Mixed Method

This method is a combination of a regression or probit, depending on the type of dependent variable, and hot-decking (by the level of the predicted outcome). The method has two steps. The first step involves estimating a probit for discrete variables, or a regression for continuous ones, based on the observations that are available. We also create a set of random numbers as the secondary sort variable to ensure the reproducibility of the variable to be imputed. In the second step we order the observations on the basis of the predicted value or predicted probability of the dependent variable estimated in step one, and the random numbers. Then we select the closest observation preceding the one with a missing value and replace the missing with that observed value.

To perform any regression or probit estimation we construct a set of covariates. These include variables indicating gender, age, race, education, marital status, wage, if employee or self-employed, industry, occupation, job tenure, firm size, union membership, interaction of gender and marital status, interaction of gender and whether the respondent was an employee or self-employed and whether (s)he was working on a full or part-time basis. In addition to these covariates we construct an indicator representing a DC plan only<sup>5</sup> for imputing DC account balances. A dummy variable indicates when an observation for a covariate is missing. We also construct a compact version of above covariates by using the age and age squared, wage and wage squared as continuous variables and a dummy variable representing the missing values for each of these two variables. We also organize the education, industry, occupation, and race variables in a much smaller number of classes. We use the compact version of covariates when the sample size is too small for the expanded version. Not all covariates are used for all imputations. The list of covariates is represented in Table 2, *Imputations for Pension-Related Variables Cross-Reference*<sup>6</sup>.

### Hot-decking (by a set of random numbers)

This method is used when the sample size is too small for a regression or probit estimation. The hot-decking procedure involves assigning a random number to each of the cases, then sorting the data by the random number in descending order, and finally replacing a missing value by the neighboring observed value. The main difference between this method and the Mixed method is that in the Mixed method we use the predicted probability, or we use the predicted value of the dependent variable, as the primary sort variable. We also use a set of random numbers as the secondary sort variable to ensure the reproducibility of the results in the future if needed.

---

<sup>4</sup> Since the imputations are cross-sectional we do not use any covariates from a previous wave. This means that when we are imputing the pension related variables from a previous job for a respondent who has changed his job since the last interview, we would not use the Mixed method. The detailed covariates describing the previous job could only come from an earlier wave. The imputation method for this set of variables is hot-decking or replacement.

<sup>5</sup> DC Plan only refers to cases with one or more DC plans. Those cases do not have any DB or Combination plan.

<sup>6</sup> The appendix to this document; see file ImpPenTbls.pdf

However, in the Hot-decking method the primary sort variable is a set of random numbers with a uniform distribution<sup>7</sup>. There is no need for the secondary sort variable because the random numbers are unique for each observation. The seed for creating the random number is the number of the question that is being imputed.

## Replacement

This technique is used by following the design of the instrument, when the information from other related variables is available, or the sample size is too small for the mixed or hot-decking procedures.

1. The design of the instrument: when the response to the type of pension question is Don't Know and in some cases Refused, the instrument treats that unknown as if it was a type DB plan. That means respondents with a Don't Know (in all sections) and Refused (in some sections) for the "type of pension plan" variable are asked the follow-up questions for type DB. Therefore, we follow that convention and replace those Don't Knows and Refusals with type DB in Waves 1 to 4 when appropriate. In Wave 5 the Don't Know responses are replaced with type DB if there were valid observations for the disposition of type DB plan (G32) in previous pension sequence (G30-G38) or for the amount of benefits (G72 or G82) for the same or new pension sequences. Otherwise, Don't know and Refused responses are treated as missing values.
2. By the way of deduction; if there was an observed value for the account balances for plan type DC but the plan type value was missing the missing plan type variables is replaced with type DC.
3. When the sample size is too small for mixed or hot-decking the missing value is replaced with the observed. That is true for when there is only one observed value. For example, where there are two cases with DC plans and one has a valid observed value and the other has a missing value we replace the missing value with the observed value.

The imputed variables are listed in Table 3, *Imputations for Pension-Related Variables Cross-Reference*.

### 2b3. Modified Imputations

The Mixed or Hot-decking procedure is modified for imputing the plan type and account balances variables. The plan type variable will be imputed in two rounds. In the first round, first we construct the dependent variable in the form of a dummy variable indicating if the plan is a DC type (only) plan or else. The dummy variable is equal to 1 if the plan type is DC only. It is zero if DB type only or if a combination of DB and DC. Then we use a probit model to get the estimates of the predicted dependent variable. In the second round, we hot-deck plan type for those with a DB only or DB combined with DC. The dummy variable is equal to 1 if DB only,

---

<sup>7</sup> To avoid hot-decking for several missing values in a row we tried a normal distribution for the hot-decking step. The results were not significantly different from hot-decking with a uniform distribution.

and zero if DB combined with DC. Then we use the constructed variable as the dependent variable in a probit model to estimate the probability of having a DB plan or else. Finally we hot-deck for the DB plan.

In the end we have two new imputed variables; one imputed in the first round and the other in the second round. The missing plan types are replaced by the imputed value from the first round if it is a DC type only. If the imputed value from the first round is not a DC type only the missing will be replaced by a plan type DB or Both imputed in the second round. This procedure is repeated for each of the three plans in waves 1 to 4 and for the four plans in wave 5.

The DC account balance variable from a previous or last job is imputed in two steps. Respondents with a pension plan from a previous or last job who report having a DC plan are asked both about their account balances when left and about their current balances. For those cases who did not report current balances when left, but did report a current balance, we use the current balance variable as the covariate and impute for the missing values. We follow the analogous procedure where balance when left is missing but current balance is reported. Then we use other covariates or hot-deck, depending on the sample size, for those cases where neither balance is reported.

The imputation procedure for account balances for DC plans and DC part of Both plans that have brackets in Wave 5 is modified also. The imputation process involves three steps. In the first step, we construct a variable *class* by organizing the observed continuous and bracket values of the balances variable based on the upper and lower limits of bracket variables. Then we impute for cases with complete bracket information using either mixed or hot-decking method, and use the class variable as the primary sort variable. In the second step we impute for cases with an open end limit brackets. We first, organize the observed values of the balances, including the imputed values in the first round, based on the information from bracket variables with an open upper or lower limit. Then we construct a new class variable and impute for those cases and use the new constructed variable class as the primary sort variable. In these two rounds of imputations the donors are among observations within the same class. In the third step we impute for cases with missing values (without any other information) using either mixed or hot-decking method, including the imputed variables in the first and second rounds.

## **2b4. Imputation Type Variables**

The imputation type variables indicate the type of imputation. For a single variable, different cases may have been imputed using different methods. For example while some cases may have replacement others have been Hot-decked or imputed by Mixed Method. The imputation type variables are coded by 1 for “Mixed Method”, 2 for “Hot-decking”, and 3 for any “Replacement or Adjustment”. If the observed value stays unchanged the code will be 4 for “Original values”.

## **3. Output Files**

### **3a. File Naming Conventions**

The names of the files start with “IP” (Imputed Pension), followed by the interview year and then a section indicator. The following extensions are used for types of files that are distributed:

.DA data files  
 .SAS SAS program statements  
 .SPS SPSS program statements  
 .DO Stata DO statements  
 .DCT Stata dictionary statements  
 .TXT codebook

One of these file types is provided for each of the 6 data files. For example,

IP2000G.DA contains data derived from HRS 2000 Section G,  
 IP2000G.SAS contains corresponding SAS program statements,  
 IP2000G.SPS contains corresponding SPSS program statements,  
 IP2000G.DO contains corresponding Stata do statements,  
 IP2000G.DCT contains corresponding Stata dictionary statements, and  
 IP2000G.TXT contains the ASCII codebook.

### 3b. Data Files

The output files are organized so that they match the original data files. The data files are provided in ASCII format, with fixed-length records. Use associated SAS, SPSS or State program statements to read the data into the analysis package of your choice. The files are packaged for download from our Web site in ZIP format. The distribution file, *ImpPen.zip*, contains six sets of year-level files plus this document.

Wave/Sections	SAS	SPSS	Stata	Data	Codebook
HRS 1992 (Wave 1) F, G, H	IP1992F.sas IP1992G.sas IP1992H.sas	IP1992F.sps IP1992G.sps IP1992H.sps	IP1992F.do/.dct IP1992G.do/.dct IP1992H.do/.dct	IP1992F.da IP1992G.da IP1992H.da	IP1992F.txt IP1992G.txt IP1992H.txt
HRS 1994 (Wave 2) FA, FB, FC, G, H	IP1994FA.sas IP1994FB.sas IP1994FC.sas IP1994G.sas IP1994H.sas	IP1994FA.sps IP1994FB.sps IP1994FC.sps IP1994G.sps IP1994H.sps	IP1994FA.do/.dct IP1994FB.do/.dct IP1994FC.do/.C.dct IP1994G.do/.dct IP1994H.do/.dct	IP1994FA.da IP1994FB.da IP1994FC.da IP1994G.da IP1994H.da	IP1994FA.txt IP1994FB.txt IP1994FC.txt IP1994G.txt IP1994H.txt
HRS 1996 (Wave 3) G, GG, H	IP1996G.sas IP1996GG.sas IP1996GH.sas	IP1996G.sps IP1996GG.sps IP1996GH.sps	IP1996G.do/.dct IP1996GG.do/.dct IP1996GH.do/.dct	IP1996G.da IP1996GG.da IP1996GH.da	IP1996G.txt IP1996GG.txt IP1996GH.txt
HRS 1998 (Wave 4) G, GG, H	IP1998G.sas IP1998GG.sas IP1998GH.sas	IP1998G.sps IP1998GG.sps IP1998GH.sps	IP1998G.do/.dct IP1998GG.do/.dct IP1998GH.do/.dct	IP1998G.da IP1998GG.da IP1998GH.da	IP1998G.txt IP1998GG.txt IP1998GH.txt
HRS 2000 (Wave 5) G, GG, H	IP2000G.sas IP2000GG.sas IP2000GH.sas	IP2000G.sps IP2000GG.sps IP2000GH.sps	IP2000G.do/.dct IP2000GG.do/.dct IP2000GH.do/.dct	IP2000G.da IP2000GG.da IP2000GH.da	IP2000G.txt IP2000GG.txt IP2000GH.txt
HRS 2002 (Wave 6) J, K, L	IP2002J.sas IP2002K.sas IP2002L.sas	IP2002J.sps IP2002K.sps IP2002L.sps	IP2002J.do/.dct IP2002K.do/.dct IP2002L.do/.dct	IP2002J.da IP2002K.da IP2002L.da	IP2002J.txt IP2002K.txt IP2002L.txt

### 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:\ipen	File downloaded from Web site
c:\ipen\codebook	codebook and documentation files
c:\ipen\data	data files
c:\ipen\sas	SAS command files
c:\ipen\spss	SPSS command files
c:\ipen\stata	Stata command files

Decompress the .zip file into the appropriate subdirectories. You will need approximately 42 MB of free space on your storage device to store these files.

## 4. Identification Variables

Identification variables for the *Imputations for Pension-Related Variables* files are stored in character format. Several variables, HHID, PN, and xSUBHH are used to uniquely identify the nine different level data sets that comprise this data release.

### 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.

### 4c. xSUBHH – Sub-household Identifier

In combination with HHID, xSUBHH uniquely identifies a household for a given wage. Sub-household identifiers can be different at each wave. xSUBHH has one digit..

## 5. 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 extension, those with SAS statements with a .SAS extension, and those with Stata statements with .DO and .DCT extensions.

The statement files are named beginning with the same prefix as the corresponding data file. For example, SAS statements in the file CF2002\_G.SAS go with the CF2002\_G.DA data file.

### **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:\IPen\sas" and the data file is located in "c:\IPen\data", you can run the file as is. A SAS system file (\*.SD2 or \*.SAS7BDAT) will be saved to directory "c:\IPen\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:\IPen\spss" and the data file is located in "c:\IPen\data", you can run the file as is. An SPSS system file (\*.SAV) will be saved to directory "c:\IPen\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:\IPen\stata" and the data file is located in "c:\IPen\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” at 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 [hrrquest@isr.umich.edu](mailto:hrrquest@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 to using the *Imputations for Pension-Related Variables* 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.

### **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: [hrsquest@isr.umich.edu](mailto:hrsquest@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